

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



458 459

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12Q 1/70, 1/68, C12N 15/10, 15/34, 1/21, C07K 14/01, C12Q 1/18		A2	(11) International Publication Number: WO 00/32825 (43) International Publication Date: 8 June 2000 (08.06.00)																		
(21) International Application Number: PCT/IB99/02040 (22) International Filing Date: 3 December 1999 (03.12.99) (30) Priority Data: <table border="0"><tr><td>60/110,992</td><td>3 December 1998 (03.12.98)</td><td>US</td></tr><tr><td>09/326,144</td><td>3 June 1999 (03.06.99)</td><td>US</td></tr><tr><td>09/407,804</td><td>28 September 1999 (28.09.99)</td><td>US</td></tr><tr><td>60/157,218</td><td>30 September 1999 (30.09.99)</td><td>US</td></tr><tr><td>60/168,777</td><td>1 December 1999 (01.12.99)</td><td>US</td></tr><tr><td>09/454,252</td><td>2 December 1999 (02.12.99)</td><td>US</td></tr></table> (71) Applicant (for all designated States except US): PHAGETECH, INC. [CA/CA]; Place du Parc, Case Postale 387, Montreal H2W 2N9 (CA). (72) Inventors; and (75) Inventors/Applicants (for US only): PELLETIER, Jerry [CA/CA]; 8 Lakeview, Baie D'Urfe, Quebec H9X 3B1 (CA). GROS, Phillippe [CA/CA]; 107 Montrose, St. Lambert, Quebec J4R 1X4 (CA). DUBOW, Michael [CA/CA]; 4901 Coolbrook Avenue, Montreal, Quebec H3X 2K8 (CA).			60/110,992	3 December 1998 (03.12.98)	US	09/326,144	3 June 1999 (03.06.99)	US	09/407,804	28 September 1999 (28.09.99)	US	60/157,218	30 September 1999 (30.09.99)	US	60/168,777	1 December 1999 (01.12.99)	US	09/454,252	2 December 1999 (02.12.99)	US	(74) Agents: MORROW, Joy, D. et al.; Smart & Biggar, 900 - 55 Metcalfe Street, P.O. Box 2999, Station D, Ottawa, Ontario K1P 5Y6 (CA). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
60/110,992	3 December 1998 (03.12.98)	US																			
09/326,144	3 June 1999 (03.06.99)	US																			
09/407,804	28 September 1999 (28.09.99)	US																			
60/157,218	30 September 1999 (30.09.99)	US																			
60/168,777	1 December 1999 (01.12.99)	US																			
09/454,252	2 December 1999 (02.12.99)	US																			
(54) Title: DEVELOPMENT OF NOVEL ANTI-MICROBIAL AGENTS BASED ON BACTERIOPHAGE GENOMICS																					
(57) Abstract <p>A method for identifying suitable targets for antibacterial agents based on identifying targets of bacteriophage-encoded proteins is described. Also described are compositions useful in the identification methods and in inhibiting bacterial growth, and methods for preparing and using such compositions.</p>																					

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

DESCRIPTION

Development of Novel Anti-Microbial Agents Based on Bacteriophage Genomics

5 BACKGROUND OF THE INVENTION

The present invention relates to the field of antibacterial agents and the treatment of infections of animals or other complex organisms by bacteria.

10 The frequency and spectrum of antibiotic-resistant infections have, in recent years, increased in both the hospital and community. Certain infections have become essentially untreatable and are growing to epidemic proportions in the developing world as well as in institutional settings in the developed world. The staggering spread of antibiotic resistance in pathogenic bacteria has been attributed to microbial
15 genetic characteristics, widespread use of antibiotic drugs, and changes in society that enhance the transmission of drug-resistant organisms. This spread of drug resistant microbes is leading to ever increasing morbidity, mortality and health-care costs.

 Ironically, it is the very success of antibiotics, resulting in their widespread use, that has contributed the most to rising numbers of drug resistant bacterial strains.
20 The longer a bacterial strain is exposed to a drug, the more likely it is to acquire resistance. Today, a total of 160 antibiotics, all based on a few basic chemical structures and targeting a small number of metabolic pathways, have found their way to market. Over-prescription of these drugs, as well as the failure of patients to comply with the complete antibiotic regimen, has lead to the rapid emergence of
25 antibiotic resistant strains. Such misuse of prescriptions, careless use of antibiotics in virtually all commercial production of beef and fowl, and changing societal conditions, such as the growth of day-care centers, increased long-term care in hospitals, and increased mobility of the population, has provided an environment where drug-resistant microbes can emerge and spread. Thus, virtually all common
30 infectious bacteria are becoming, or have already become, resistant to one or more groups of antibiotics. Such resistance now reaches all classes of antibiotics currently in use, including: β -lactams, fluoroquinolones, aminoglycosides, macrolide peptides, chloramphenicol, tetracyclines, rifampicin, folate inhibitors, glycopeptides, and
 mupirocin.

35 Over the last 45 years bacteria have adapted genetically to avoid the destruction/alteration of the essential pathways that these chemotherapeutic agents

target. Antibiotic resistant bacterial strains are now emerging at a higher rate than the rate at which new antibiotics are being developed. The consequence of this dilemma has been a dramatic increase in the cost of treating infections what would otherwise easily succumb to routine antibiotic therapy. Furthermore, and perhaps most
5 importantly, the emergence of multiple drug resistant pathogenic bacteria has led to a significant increase in morbidity and mortality, particularly in institutional settings.

Most major pharmaceutical companies have on-going drug discovery programs for novel anti-microbials. These are based on screens for small molecule inhibitors (natural products, bacterial culture media, libraries of small molecules,
10 combinatorial chemistry) of crucial metabolic pathways of the micro-organism of interest (*e.g.*, bacteria, fungi, parasites, worms). The screening process is largely for cytotoxic compounds and in most cases is not based on a known mechanism of action of the compounds. Pharmaceutical companies have large programs in this area. Classical drug screening programs are being exhausted and many of these
15 pharmaceutical companies are looking towards rational drug design programs.

Several small to mid-size biotechnology companies as well as large pharmaceutical companies have developed systematic high-throughput sequencing programs to decipher the genetic code of specific micro-organisms of interest. The goal is to identify, through sequencing, unique biochemical pathways or intermediates
20 that are unique to the microorganism. Knowledge of this may, in turn, form the rationale for a drug discovery program based on the mechanism of action of the identified enzymes/proteins. Genome Therapeutics Corp., The Institute for Genome Research, Human Genome Sciences Inc., and other companies have such sequencing programs in place. However, one of the most critical steps in this approach is the
25 ascertainment that the identified proteins and biochemical pathways are 1) non-redundant and essential for bacterial survival, and 2) constitute suitable and accessible targets for drug discovery.

SUMMARY OF THE INVENTION

While animals such as humans are, on occasion, infected by pathogenic bacteria, bacteria also have natural enemies. A number of host-specific viruses, known as bacteriophages or phages, infect and kill bacteria in the natural environment. Such bacteriophages generally have small compact genomes and bacteria are their exclusive hosts. Many known bacteria are host to a large number of bacteriophages that have been described in the literature. During the 1940's - 1960's, phage biology was an area of active research. As a testimony to this, the study of phages which infect and inhibit the enteric bacterium *Escherichia coli* (*E. coli*) contributed much to the early understanding of molecular biology and virology.

As is generally understood, bacteriophage (or phages) are viruses that infect and kill bacteria. They are natural enemies of bacteria and, over the course of evolution, have developed proteins (products of DNA sequences) which enable them to infect a host bacteria, replicate their genetic material, usurp host metabolism, and ultimately kill their host. The scientific literature well documents the fact that many known bacteria have a large number of such bacteriophages (Ackermann and DuBow, 1987) that can infect and kill them (for example, see the ATCC bacteriophage collection at <http://www.atcc.org>).

This invention utilizes the observation that bacteriophages successfully infect and inhibit or kill host bacteria, targeting a variety of normal host metabolic and physiological traits, some of which are shared by all bacteria, pathogenic and nonpathogenic alike. The term "pathogenic" as used herein denotes a contribution to or implication in disease or a morbid state of an infected organism. The invention thus involves identifying and elucidating the molecular mechanisms by which phages interfere with host bacterial metabolism, an objective being to provide novel targets for drug design. Whether the phage blocks bacterial RNA transcription or translation, or attacks other important metabolic pathways, such as cell wall assembly or membrane integrity, the basic blueprint for a phage's bacteria-inhibiting ability is encoded in its genome and can be unlocked using bioinformatics, functional genomics, and proteomics. By these means, the invention utilizes sequence information from the genomics of bacteriophage to identify novel antimicrobials that can be further used to actively and/or prophylactically treat bacterial infection.

Two important components of the invention thus are: i) the identification of bacteria-inhibiting phage open reading frames ("ORF"s) and corresponding products that can be used to develop antibiotics based on amino acid sequence and secondary structural characteristics of the ORF products, and ii) the use of bacteriophages to map

out essential bacterial target genes and homologs, which can in turn lead to the development of suitable anti-microbial agents. These two avenues represent new and general methods for developing novel antimicrobials.

The invention thus concerns the identification of bacteriophage ORFs that supply bacteria-inhibiting functions. In this regard, use of the terms "inhibit", "inhibition", "inhibitory", and "inhibitor" all refer to a function of reducing a biological activity or function. Such reduction in activity or function can, for example, be in connection with a cellular component, *e.g.*, an enzyme, or in connection with a cellular process, *e.g.*, synthesis of a particular protein, or in connection with an overall process of a cell, *e.g.*, cell growth. In reference to bacterial cell growth, for example, an inhibitory effect (*i.e.*, a bacteria-inhibiting effect) may be bacteriocidal (killing of bacterial cells) or bacteriostatic (*i.e.*, stopping or at least slowing bacterial cell growth). The latter slows or prevents cell growth such that fewer cells of the strain are produced relative to uninhibited cells over a given period of time. From a molecular standpoint, such inhibition may equate with a reduction in the level of, or elimination of, the transcription and/or translation of a specific bacterial target(s), or reduction or elimination of activity of a particular target biomolecule.

It is particularly advantageous to evaluate a plurality of different phage ORFs for inhibitory activity that may be from one, but is preferably from a plurality of different phage. For example, evaluating ORFs from a number of different phage of the same bacterial host provides at least two advantages. One is that the multiple phages will provide identification of a variety of different targets. Second, it is likely that multiple phage will utilize the same cellular target

As used herein, the terms "bacteriophage" and "phage" are used interchangeably to refer to a virus which can infect a bacterial strain or a number of different bacterial strains.

In the context of this invention, the term "bacteriophage ORF" or "phage ORF" or similar term refers to a nucleotide sequence in or from a bacteriophage. In connection with a particular ORF, the terms refer an open reading frame which has at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity with an ORF from the particular phage identified herein (*e.g.*, with an ORF as identified herein) or to a nucleic acid sequence which has the specified sequence identity percentage with such an ORF sequence.

A first aspect of the invention thus provides a method for identifying a bacteriophage nucleic acid coding region encoding a product active on an essential bacterial target by identifying a nucleic acid sequence encoding a gene product which

provides a bacteria-inhibiting function when the bacteriophage infects a host bacterium, preferably one that is an animal or plant pathogen, more preferably a bird or mammalian pathogen, and most preferably a human pathogen. The bacteriophage is an uncharacterized bacteriophage. Thus, the method excludes, for example, phage λ , ϕ x174, m13 and other *E.coli*-specific bacteriophage that have been studied with respect to gene number and/or function. It also excludes, for example, the nucleic acid coding regions described in Tables 12-14, and in preferred embodiments, excludes the phage in which those regions are naturally located.

In connection with bacteriophage, the term "uncharacterized" means that a certain bacteriophage's genome has not yet been fully identified such that the genes having function involved in inhibiting host cells have not been identified. In particular, phage for which the description of genomic or protein sequence was first provided herein are uncharacterized. Phage sequences for which host bacteria-inhibiting functions have been identified prior to the filing of the present application (or alternatively prior to the present invention) are specifically excluded from the aspects involving utilization of sequences from uncharacterized bacteriophage, except that aspects may involve a plurality of phage where one or more of those phage are uncharacterized and one or more others have been characterized to some extent. A number of different bacteria-inhibiting phage ORFs are indicated in Tables 11-14. The phage ORFs or sequences identified therein are not within the term "uncharacterized; alternatively, in preferred embodiments the phage containing those ORFs are excluded from this term. Further, any additional phage ORFs (or alternatively the phage which contain those ORFs) which have previously been described in the art as bacteria-inhibiting ORFs are expressly excluded; those ORFs or phage are known to those skilled in the art and the exclusion can be made express by specifically naming such ORFs or phage as needed (likewise for uncharacterized targets as described below). For the sake of brevity, such a listing is not expressly presented, as such information is readily available to those skilled in the art.

Stating that an agent or compound is "active on" a particular cellular target, such as the product of a particular gene, means that the target is an important part of a cellular pathway which includes that target and that the agent acts on that pathway. Thus, in some cases the agent may act on a component upstream or downstream of the stated target, including on a regulator of that pathway or a component of that pathway.

By "essential", in connection with a gene or gene product, is meant that the host cannot survive without, or is significantly growth compromised, in the absence of depletion, or alteration of functional product. An "essential gene" is thus one that encodes a product that is beneficial, or preferably necessary, for cellular growth in

vitro in a medium appropriate for growth of a strain having a wild-type allele corresponding to the particular gene in question. Therefore, if an essential gene is inactivated or inhibited, that cell will grow significantly more slowly, preferably less than 20%, more preferably less than 10%, most preferably less than 5% of the growth rate of the uninhibited wild-type, or not at all, in the growth medium. Preferably, in the absence of activity provided by a product of the gene, the cell will not grow at all or will be non-viable, at least under culture conditions similar to the *in vivo* conditions normally encountered by the bacterial cell during an infection. For example, absence of the biological activity of certain enzymes involved in bacterial cell wall synthesis can result in the lysis of cells under normal osmotic conditions, even though protoplasts can be maintained under controlled osmotic conditions. In the context of the invention, essential genes are generally the preferred targets of antimicrobial agents. Essential genes can encode target molecules directly or can encode a product involved in the production, modification, or maintenance of a target molecule.

A "target" refers to a biomolecule that can be acted on by an exogenous agent, thereby modulating, preferably inhibiting, growth or viability of a cell. In most cases such a target will be a nucleic acid sequence or molecule, or a polypeptide or protein. However, other types of biomolecules can also be targets, *e.g.*, membrane lipids and cell wall structural components.

The term "bacterium" refers to a single bacterial strain, and includes a single cell, and a plurality or population of cells of that strain unless clearly indicated to the contrary. In reference to bacteria or bacteriophage, the term "strain" refers to bacteria or phage having a particular genetic content. The genetic content includes genomic content as well as recombinant vectors. Thus, for example, two otherwise identical bacterial cells would represent different strains if each contained a vector, *e.g.*, a plasmid, with different phage ORF inserts.

In preferred embodiments, the phage is *Staphylococcus aureus* phage 77, 3A, 96, or 44 AHJD, *Enterococcus* sp. phage 182, or *Streptococcus pneumoniae* phage Dp-1.

In preferred embodiments, the phage is selected from. Preferred embodiments involve expressing at least one recombinant phage ORF(s) in a bacterial host followed by inhibition analysis of that host. Inhibition following expression of the phage ORF is indicative that the product of the ORF is active on an essential bacterial target. Such evaluation can be carried out in a variety of different formats, such as on a support matrix such as a solidified medium in a petri dish, or in liquid culture.

Preferably a plurality of phage ORFs are expressed in at least one bacterium. The plurality of phage ORFs can be from one or a plurality of phage. With respect to a single phage or at least one phage in a plurality of phages, the plurality of expressed ORFs preferably represents at least 10%, more preferably at least 20%, 40%, or 60%, still more preferably at least 80% or 90%, and most preferably at least 95% of the ORFs in the phage genome. Preferably, for a plurality of phage, the plurality of expressed ORFs preferably represents at least 10%, more preferably at least 20%, 40%, or 60%, still more preferably at least 80% or 90%, and most preferably at least 95% of the ORFs in the phage genome of each phage. The plurality of phage ORFs can be expressed in a single bacterium, or in a plurality of bacteria where one ORF is expressed in each bacterium, or in a plurality of bacteria where a plurality of ORFs are expressed in at least one or in all of the plurality of bacteria, or combinations of these.

In embodiments of the above aspect (as well as in other aspects herein) in which a plurality of phage are utilized, a plurality of phage have the same bacterial host species; have different bacterial host species; or both. The plurality of phage includes at least two different phage, preferably at least 3, 4, 5, 6, 8, 10, 15, 20, or more different phage. Indeed, more preferably, the plurality of phage will include 50, 75, 100, or more phage. As described herein, the larger number of phage is useful to provide additional target and target evaluation information useful in developing antibacterial agents, for example, by providing identification of a larger range of bacterial targets, and/or providing further indication of the suitability of a particular target (for example, utilization of a target by a number of different unrelated phage can suggest that the target is particularly stable and accessible and effective) and/or can indicate alternate sites on a target which interact with different inhibitors.

Further embodiments involve confirmation of the inhibitor function of the phage ORF, such as by utilizing or incorporating a control(s) designed to confirm the inhibitory nature of the ORF(s) being evaluated. The control can, for example, be provided by expression of an inactive or partially inactive form of the ORF or ORF product, and/or by the absence of expression of the ORF or ORF product in the same or a closely comparable bacterial strain as that used for expression of the test ORF. The reduced level of activity or the absence of active ORF product in the control will thus not provide the inhibition provided by a corresponding inhibitory ORF, or will provide a distinguishably lower level of inhibition. An inactivated or partially inactivated control has a mutation(s), e.g., in the coding region or in flanking regulatory elements, that reduce(s) or eliminate(s) the normal function of the ORF. Thus, the inhibition of a bacterium following expression of a phage ORF is determined by comparison with the effects of expression of an inactivated ORF or the

response of the bacteria in the absence of expression in the same or similar type bacterium. Such determination of inhibition of the bacterium following expression of the ORF is indicative of a bacteria-inhibiting function. These manipulations are routinely understood and accomplished by those of skill in the art using standard techniques. In embodiments utilizing absence of expression of the ORF, the bacteria can, for example, contain an empty vector or a vector which allows expression of an unrelated sequence which is preferably non-inhibitory. Alternatively, the bacteria may have no vector at all. Combinations of such controls or other controls may also be utilized as recognized by those skilled in the art.

10 In embodiments involving expression of a phage ORF in a bacterial strain, in preferred embodiments that expression is inducible.

By "inducible" is meant that expression is absent or occurs at a low level until the occurrence of an appropriate environmental stimulus provides otherwise. For the present invention such induction is preferably controlled by an artificial environmental change, such as by contacting a bacterial strain population with an inducing compound (*i.e.*, an inducer). However, induction could also occur, for example, in response to build-up of a compound produced by the bacteria in the bacterial culture, *e.g.*, in the medium. As uncontrolled or constitutive expression of inhibitory ORFs can severely compromise bacteria to the point of eradication, such expression is therefore undesirable in many cases because it would prevent effective evaluation of the strain and inhibitor being studied. For example, such uncontrolled expression could prevent any growth of the strain following insertion of a recombinant ORF, thus preventing determination of effective transfection or transformation. A controlled or inducible expression is therefore advantageous and is generally provided through the provision of suitable regulatory elements, *e.g.*, promoter/operator sequences that can be conveniently transcriptionally linked to a coding sequence to be evaluated. In most cases, the vector will also contain sequences suitable for efficient replication of the vector in the same or different host cells and/or sequences allowing selection of cells containing the vector, *i.e.*, "selectable markers." Further, preferred vectors include convenient primer sequences flanking the cloning region from which PCR and/or sequencing may be performed.

As knowledge of the nucleotide sequence of phage ORFs is useful, *e.g.*, for assisting in the identification of phage proteins active against essential bacterial host targets, preferred embodiments involve the sequencing of at least a portion of the phage genome in combination with the above methods. This can be done either before or after or independent of expression and inhibition of the ORF in the bacteria, and provides information on the nature and characteristics of the ORF. Such a portion is

preferably at least 10%, 20%, 40%, 80%, 90%, or 100% of the phage genome. For embodiments in which a plurality of phage are utilized, preferably each phage is sequenced to an extent as just specified.

- Such sequencing is preferably accompanied by computer sequence analysis to
- 5 define and evaluate ORF(s), ORF products, structural motifs or functional properties of ORF products, and/or their genetic control elements. Thus, certain embodiments incorporate computer sequence analyses or nucleic acid and/or amino acid sequences. Further, existing data banks can provide phage sequence and product information which can be utilized for analysis and identification of ORFs in the sequence.
- 10 Computer analysis may further employ known homologous sequences from other species that suggest or indicate conserved underlying biochemical function(s) for the inhibitory or potentially inhibitory ORF sequence(s) being evaluated. This can include the sequences of signature motifs of identified classes of inhibitors.

- In the context of the phage nucleic acid sequences, e.g., gene sequences, of this
- 15 invention, the terms "homolog" and "homologous" denote nucleotide sequences from different bacteria or phage strains or species or from other types of organisms that have significantly related nucleotide sequences, and consequently significantly related encoded gene products, preferably having related function. Homologous gene sequences or coding sequences have at least 70% sequence identity (as defined by the
- 20 maximal base match in a computer-generated alignment of two or more nucleic acid sequences) over at least one sequence window of 48 nucleotides, more preferably at least 80 or 85%, still more preferably at least 90%, and most preferably at least 95%. The polypeptide products of homologous genes have at least 35% amino acid sequence identity over at least one sequence window of 18 amino acid residues, more
- 25 preferably at least 40%, still more preferably at least 50% or 60%, and most preferably at least 70%, 80%, or 90%. Preferably, the homologous gene product is also a functional homolog, meaning that the homolog will functionally complement one or more biological activities of the product being compared. For nucleotide or amino acid sequence comparisons where a homology is defined by a % sequence
- 30 identity, the percentage is determined using BLAST programs (with default parameters (Altschul et al., 1997, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acid Res.* 25:3389-3402). Any of a variety of algorithms known in the art which provide comparable results can also be used, preferably using default parameters. Performance characteristics for
- 35 three different algorithms in homology searching is described in Salamov et al., 1999, "Combining sensitive database searches with multiple intermediates to detect distant

homologues." *Protein Eng.* 12:95-100. Another exemplary program package is the GCG™ package from the University of Wisconsin.

Homologs may also or in addition be characterized by the ability of two complementary nucleic acid strands to hybridize to each other under appropriately stringent conditions. Hybridizations are typically and preferably conducted with probe-length nucleic acid molecules, preferably 20-100 nucleotides in length. Those skilled in the art understand how to estimate and adjust the stringency of hybridization conditions such that sequences having at least a desired level of complementarity will stably hybridize, while those having lower complementarity will not. For examples of hybridization conditions and parameters, see, *e.g.*, Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J. Homologs and homologous gene sequences may thus be identified using any nucleic acid sequence of interest, including the phage ORFs and bacterial target genes of the present invention.

A typical hybridization, for example, utilizes, besides the labeled probe of interest, a salt solution such as 6xSSC (NaCl and Sodium Citrate base) to stabilize nucleic acid strand interaction, a mild detergent such as 0.5% SDS, together with other typical additives such as Denhardt's solution and salmon sperm DNA. The solution is added to the immobilized sequence to be probed and incubated at suitable temperatures to preferably permit specific binding while minimizing nonspecific binding. The temperature of the incubations and ensuing washes is critical to the success and clarity of the hybridization. Stringent conditions employ relatively higher temperatures, lower salt concentrations, and/or more detergent than do non-stringent conditions. Hybridization temperatures also depend on the length, complementarity level, and nature (ie, "GC content") of the sequences to be tested. Typical stringent hybridizations and washes are conducted at temperatures of at least 40°C, while lower stringency hybridizations and washes are typically conducted at 37°C down to room temperature (~25°C). One of skill in the art is aware that these conditions may vary according to the parameters indicated above, and that certain additives such as formamide and dextran sulphate may also be added to affect the conditions.

By "stringent hybridization conditions" is meant hybridization conditions at least as stringent as the following: hybridization in 50% formamide, 5X SSC, 50 mM NaH₂PO₄, pH 6.8, 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5X Denhart's solution at 42°C overnight; washing with 2X SSC, 0.1% SDS at 45°C; and washing with 0.2X SSC, 0.1% SDS at 45°C.

In sequence comparison analyses, an ORF, or motif, or set of motifs in a bacteriophage sequence can be compared to known inhibitor sequences, *e.g.*, homologous sequences encoding homologous inhibitors of bacterial function. Likewise, the analysis can include comparison with the structure of essential bacterial gene products, as structural similarities can be indicative of similar or replacement biological function. Such analysis can include the identification of a signature, or characteristic motif(s) of an inhibitor or inhibitor class.

Also, the identification of structural motifs in an encoded product, based on nucleotide or amino acid sequence analysis, can be used to infer a biochemical function for the product. A database containing identified structural motifs in a large number of sequences is available for identification of motifs in phage sequences. The database is PROSITE, which is available at www.expasy.ch/cgi-bin/scanprosite. The identification of motifs can, for example, include the identification of signature motifs for a class or classes of inhibitory proteins. Other such databases may also be used.

In aspects and preferred embodiments described herein, in which a bacterium or host bacterium is specified, the bacterium or host bacterium is preferably selected from a pathogenic bacterial species, for example, one selected from Table 1. Preferably, an animal or plant pathogen is used. For animals, preferably the bacterium is a bird or mammalian pathogen, still more preferably a human pathogen.

In aspects and preferred embodiments involving a bacteriophage or sequences from a bacteriophage, one or more bacteriophage are preferably selected from those listed in Table 1. Those exemplary bacteriophage are readily obtained from the indicated sources.

In some cases, it is advantageous to utilize phage with non-pathogenic host bacteria. The genome, structural motif, ORF, homolog, and other analyses described herein can be performed on such phage and bacteria. Such analysis provides useful information and compositions. The results of such analyses can also be utilized in aspects of the present invention to identify homologous ORFs, especially inhibitor ORFs in phage with pathogenic bacterial hosts. Similarly, identification of a target in a non-pathogenic host can be used to identify homologous sequences and targets in pathogenic bacteria, especially in genetically closely related bacteria. Those skilled in the art are familiar with bacterial genetic relationships and with how to determine relatedness based on levels of genomic identity or other measures of nucleotide sequence and/or amino acid sequence similarity, and/or other physical and culture characteristics such as morphology, nutritional requirements, or minimal media-to support growth.

Also in preferred embodiments, an embodiment of this aspect is combined with an embodiment of the following aspect.

A related aspect of the invention provides methods for identifying a target for antibacterial agents by identifying the bacterial target(s) of at least one uncharacterized or untargeted inhibitor protein or RNA from a bacteriophage. Such identification allows the development of antibacterial agents active on such targets. Preferred embodiments for identifying such targets involve the identification of binding of target and phage ORF products to one another. The phage ORF products may be subportions of a larger ORF product that also binds the host target. In preferred embodiments, the phage protein or RNA is from an uncharacterized bacteriophage in Table 1. This aspect preferably includes the identification of a plurality of such targets in one or a plurality of different bacteria, preferably in one or a plurality of bacteria listed in Table 1.

In preferred embodiments of this aspect and other aspects of this invention involving particular phage ORFs or phage sequences, the ORF is *Staphylococcus aureus* phage 77 ORF 17, 19, 43, 102, 104, or 182 as identified in U.S. application 09/407,804, *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

As indicated for the above aspect, preferably the method involves the use of a plurality of different phage, and thus a plurality of different phage inhibitors and/or inhibitor ORFs.

In addition to uncharacterized phage ORF products, it is also useful to identify the targets of phage ORF products which are known to be inhibitors of host bacteria, but where the target has not been identified. Thus, such inhibitors can likewise be utilized as "untargeted" inhibitor phage ORFs and ORF products, e.g., proteins or RNAs.

In the context of inhibitor proteins or RNAs from a phage, the term "uncharacterized" means that a bacteria-inhibiting function for the protein has not previously been identified. Preferably, but not necessarily, the sequence of the protein or the corresponding coding region or ORF was not described in the art before the filing of the present application for patent (or alternatively prior to the present invention). Thus, this term specifically excludes any bacteria-inhibiting phage protein and its associated bacterial target which has been identified as inhibitory before the present invention or alternatively before the filing of the present application, for example those identified in Tables 12-14 or otherwise identified herein. For example, from *E. coli*, phage T7 genes 0.7 and 2.0 target the host RNA polymerase, phage T4

gp55/gp33 alter the specificity of host RNA polymerase. The T4 *regB* gene product also targets the host translation apparatus. As with the uncharacterized bacteriophage ORFs or bacteriophage above, for such identified proteins, the sequences encoding those proteins are excluded from the uncharacterized inhibitor proteins.

- 5 The term "fragment" refers to a portion of a larger molecule or assembly. For proteins, the term "fragment" refers to a molecule which includes at least 5 contiguous amino acids from the reference polypeptide or protein, preferably at least 8, 10, 12, 15, 20, 30, 50 or more contiguous amino acids. In connection with oligo- or polynucleotides, the term "fragment" refers to a molecule which includes at least 15
10 contiguous nucleotides from a reference polynucleotide, preferably at least 24, 30, 36, 45, 60, 90, 150, or more contiguous nucleotides.

- Preferred embodiments involve identification of binding that include methods for distinguishing bound molecules, for example, affinity chromatography, immunoprecipitation, crosslinking, and/or genetic screen methods that permit
15 protein:protein interactions to be monitored. One of skill in the art is familiar with these techniques and common materials utilized (see, e.g., Coligan, J. et al. (eds.) (1995) Current Protocols in Protein Science, John Wiley & Sons, Secaucus, N.J.).

- Genetic screening for the identification of protein:protein interactions typically involves the co-introduction of both a chimeric bait nucleic acid sequence (here, the
20 phage ORF to be tested) and a chimeric target nucleic acid sequence that, when co-expressed and having affinity for one another in a host cell, stimulate reporter gene expression to indicate the relationship. A "positive" can thus suggest a potential inhibitory effect in bacteria. This is discussed in further detail in the Detailed Description section below. In this way, new bacterial targets can be identified that are
25 inhibited by specific phage ORF products or derivatives, fragments, mimetics, or other molecules.

- Other embodiments involve the identification and/or utilization of mutant targets by virtue of their host's relatively unresponsive nature in the presence of expression of ORFs previously identified as inhibitory to the non-mutant or wild-type
30 strain. Such mutants have the effect of protecting the host from an inhibition that would otherwise occur and indirectly allow identification of the precise responsible target for follow-up studies and anti-microbial development. In certain embodiments, rescue from inhibition occurs under conditions in which a bacterial target or mutant target is highly expressed. This is performed, for example, through coupling of the
35 sequence with regulatory element promoters, e.g., as known in the art, which regulate expression at levels higher than wild-type, e.g., at a level sufficiently higher that the

inhibitor can be competitively bound to the highly expressed target such that the bacterium is detectably less inhibited.

Identification of the bacterial target can involve identification of a phage-specific site of action. This can involve a newly identified target, or a target where the phage site of action differs from the site of action of a previously known antibacterial agent or inhibitor. For example, phage T7 genes 0.7 and 2.0 target the host RNA polymerase, which is also the cellular target for the antibacterial agent, rifampin. To the extent that a phage product is found to act at a different site than previously described inhibitors, aspects of the present invention can utilize those new, phage-specific sites for identification and use of new agents. The site of action can be identified by techniques well-known to those skilled in the art, for example, by mutational analysis, binding competition analysis, and/or other appropriate techniques.

Once a bacterial host target protein or nucleic acid or mutant target sequence has been identified and/or isolated, it too can be conveniently sequenced, sequence analyzed (e.g., by computer), and the underlying gene(s), and corresponding translated product(s) further characterized. Preferred embodiments include such analysis and identification. Preferably such a target has not previously been identified as an appropriate target for antibacterial action.

Certain embodiments include the identification of at least one inhibitory phage ORF or ORF product, e.g., as described for the above aspect, and thus are a combination of the two aspects.

Additionally, the invention provides methods for identifying targets for antibacterial agents by identifying homologs of a bacterial target e.g., *S. aureus*, *Enterococcus faecalis* or other *Enterococci*, and *Streptococcus pneumoniae* of a bacteriophage inhibitory ORF product. Such homologs may be utilized in the various aspects and embodiments described herein as described for the host *Enterococcus* sp. for bacteriophage 182.

Other aspects of the invention provide isolated, purified, or enriched specific phage nucleic acid and amino acid sequences, subsequences, and homologs thereof for phage selected from uncharacterized phage listed in Table 1, preferably from bacteriophage 77, 3A, 96, 44AHJD (*Staphylococcus aureus* host bacterium), Dp-1 (*Streptococcus pneumoniae* host), or 182 (*Enterococcus* host) or other phage listed in Table 1 for those bacteria. For example, such sequences do not include sequences identified in any of Tables 11-14. Nucleotide sequences of this aspect are at least 15 nucleotides in length, preferably at least 18, 21, 24, or 27 nucleotides in length, more preferably at least 30, 50, or 90 nucleotides in length. In certain embodiments, longer

nucleic acids are preferred, for example those of at least 120, 150, 200, 300, 600, 900 or more nucleotides. Such sequences can, for example, be amplification oligonucleotides (e.g., PCR primers), oligonucleotide probes, sequences encoding a portion or all of a phage-encoded protein, or a fragment or all of a phage-encoded protein. In preferred embodiments, the nucleic acid sequence contains a sequence which is within a length range with a lower length as specified above, and an upper length limit which is no more than 50, 60, 70, 80, or 90% of the length of the corresponding full-length ORF. The upper length limit can also be expressed in terms of the number of base pairs of the ORF (coding region). In preferred embodiments, the nucleic acid sequence is from *Staphylococcus aureus* phage 77 ORF 17, 19, 43, 102, 104, or 182 as identified in U.S. application 09/407,804, *S. aureus* phage 44 AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

As it is recognized that alternate codons will encode the same amino acid for most amino acids due to the degeneracy of the genetic code, the sequences of this aspect includes nucleic acid sequences utilizing such alternate codon usage for one or more codons of a coding sequence. For example, all four nucleic acid sequences GCT, GCC, GCA, and GCG encode the amino acid, alanine. Therefore, if for an amino acid there exists an average of three codons, a polypeptide of 100 amino acids in length will, on average, be encoded by 3^{100} , or 5×10^{47} , nucleic acid sequences. Thus, a nucleic acid sequence can be modified (e.g., a nucleic acid sequence from a phage as specified above) to form a second nucleic acid sequence encoding the same polypeptide as encoded by the first nucleic acid sequence using routine procedures and without undue experimentation. Thus, all possible nucleic acid sequences that encode the specified amino acid sequences are also fully described herein, as if all were written out in full, taking into account the codon usage, especially that preferred in the host bacterium. The alternate codon descriptions are available in common textbooks, for example, Stryer, *BIOCHEMISTRY* 3rd ed., and Lehninger, *BIOCHEMISTRY* 3rd ed., along with many others. Codon preference tables for various types of organisms are available in the literature. Sequences with alternate codons at one or more sites can also be utilized in the computer-related aspects and embodiments herein. Because of the number of sequence variations involving alternate codon usage, for the sake of brevity, individual sequences are not separately listed herein. Instead the alternate sequences are described by reference to the natural sequence with replacement of one or more (up to all e.g., up to 3, 5, 10, 15, 20, 30, 40, 50, or more) of the degenerate codons with alternate codons from the alternate codon

table (Table 6), or a modified table applicable to a particular organism that has differing codon usage, preferably with selection according to preferred codon usage for the normal host organism or a host organism in which a sequence is intended to be expressed. Those skilled in the art also understand how to alter the alternate codons to be used for expression in organisms where certain codons code differently than shown in the "universal" codon table.

For amino acid sequences or polypeptides, sequences contain at least 5 peptide-linked amino acid residues, and preferably at least 6, 7, 10, 15, 20, 30, or 40, amino acids having identical amino acid sequence as the same number of contiguous amino acid residues in a particular phage ORF product. In some cases longer sequences may be preferred, for example, those of at least 50, 60, 70, 80, or 100 amino acids in length. In preferred embodiments, the amino acid sequence contains a sequence which is within a length range with a lower length as specified above, and an upper length limit which is no more than 50, 60, 70, 80, or 90% of the length of the corresponding full-length ORF product. The upper length limit can also be expressed in terms of the number of amino acid residues of the ORF product. In preferred embodiments, the amino acid sequence or polypeptide has bacteria-inhibiting function when expressed or otherwise present in a bacterial cell which is a host for the bacteriophage from which the sequence was derived.

By "isolated" in reference to a nucleic acid is meant that a naturally occurring sequence has been removed from its normal cellular (e.g., chromosomal) environment or is synthesized in a non-natural environment (e.g., artificially synthesized). Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide chain present, but that it is essentially free (about 90-95% pure at least) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

The term "enriched" means that the specific DNA or RNA sequence constitutes a significantly higher fraction (2-5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in cells from which the sequence was originally taken. This could be caused by a person by preferential reduction in the amount of other DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two. However, it should be noted that enriched does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased.

The term "significant" is used to indicate that the level of increase is useful to the person making such an increase and an increase relative to other nucleic acids of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no DNA or RNA from other sources. The other
5 source DNA may, for example, comprise DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes from naturally occurring events, such as viral infection, or tumor type growths, in which the level of one mRNA may be naturally increased relative to other species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to
10 elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation). Instead, it represents an indication that the sequence is relatively more pure than in the natural environment
15 (compared to the natural level, this level should be at least 2-5 fold greater, *e.g.*, in terms of mg/mL). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation
20 of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation
25 of distinct cDNA clones yields an approximately 10^6 -fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

The terms "isolated", "enriched", and "purified" as respect nucleic acids,
30 above, may similarly be used to denote the relative purity and abundance of polypeptides (multimers of amino acids joined one to another by α -carboxyl: α -amino group (peptide) bonds). These, too, may be stored in, grown in, screened in, and selected from libraries using biochemical techniques familiar in the art. Such polypeptides may be natural, synthetic or chimeric and may be extracted using any of
35 a variety of methods, such as antibody immunoprecipitation, other "tagging" techniques, conventional chromatography and/or electrophoretic methods. Some of the above utilize the corresponding nucleic acid sequence.

As indicated above, aspects and embodiments of the invention are not limited to entire genes and proteins. The invention also provides and utilizes fragments and portions thereof, preferably those which are "active" in the inhibitory sense described above. Such peptides or oligopeptides and oligo or polynucleotides have preferred
5 lengths as specified above for nucleic acid and amino acid sequences from phage; corresponding recombinant constructs can be made to express the encoded same. Also included are homologous sequences and fragments thereof.

Nucleic acid sequences of the present invention can be isolated using a method similar to those described herein or other methods known to those skilled in the art.
10 In addition, such nucleic acid sequences can be chemically synthesized by well-known methods. Also, by having particular phage ORFs, e.g., the phage ORFs identified herein (e.g., anti-bacterial ORFs of the present invention, portions thereof, or oligonucleotides derived therefrom as described), other antimicrobial sequences from other bacteriophage sources can be identified and isolated using methods
15 described here or other methods, including methods utilizing nucleic acid hybridization and/or computer-based sequence alignment methods.

The invention also provides bacteriophage antimicrobial DNA segments from other phages based on nucleic acids and sequences hybridizing to the presently identified inhibitory ORF under high stringency conditions or sequences that are
20 highly homologous. The bacteriophage segment from a specific phage, e.g., an antimicrobial DNA segment, can be used to identify a related segment from another unrelated phage based on stringent conditions of hybridization or on being a homolog based on nucleic acid and/or amino acid sequence comparisons. As with identified inhibitory sequences, such homologous coding sequences and products can be used as
25 antimicrobials, to construct active portions or derivatives, to construct peptidomimetics, and to identify bacterial targets.

The nucleotide and amino acid sequences identified herein are believed to be correct, however, certain sequences may contain a small percentage of errors, e.g., 1-5%. In the event that any of the sequences have errors, the corrected sequences can be
30 readily provided by one skilled in the art using routine methods. For example, the nucleotide sequences can be confirmed or corrected by obtaining and culturing the relevant phage, and purifying phage genomic nucleic acids. A region or regions of interest can be amplified, e.g., by PCR from the appropriate genomic template, using primers based on the described sequence. The amplified regions can then be
35 sequenced using any of the available methods (e.g., a dideoxy termination method).

This can be done redundantly to provide the corrected sequence or to confirm that the described sequence is correct. Alternatively, a particular sequence or sequences can be identified and isolated as an insert or inserts in a phage genomic library and isolated, amplified, and sequenced by standard methods. Confirmation or correction of a nucleotide sequence for a phage gene provides an amino acid sequence of the encoded product by merely reading off the amino acid sequence according to the normal codon relationships and/or expressed in a standard expression system and the polypeptide product sequenced by standard techniques. The sequences described herein thus provide unique identification of the corresponding genes, coding sequences, and other sequences, allowing those sequences to be used in the various aspects of the present invention.

In other aspects, the invention provides recombinant vectors and cells harboring at least one of the phage ORFs or portion thereof, or bacterial target sequences described herein. As understood by those skilled in the art, vectors may be provided in different forms, including, for example, plasmids, cosmids, and virus-based vectors. See, *e.g.*, Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; See also, Ausubel, F.M. et al. (eds.) (1994) Current Protocols in Molecular Biology, John Wiley & Sons, Secaucus, N.J.

In preferred embodiments, the vectors will be expression vectors, preferably shuttle vectors that permit cloning, replication, and expression within bacteria. An "expression vector" is one having regulatory nucleotide sequences containing transcriptional and translational regulatory information that controls expression of the nucleotide sequence in a host cell. Preferably the vector is constructed to allow amplification from vector sequences flanking an insert locus. In certain embodiments, the expression vectors may additionally or alternatively support expression, and/or replication in animal, plant and/or yeast cells due to the presence of suitable regulatory sequences, *e.g.*, promoters, enhancers, 3' stabilizing sequences, primer sequences, etc. In preferred embodiments, the promoters are inducible and specific for the system in which expression is desired, *e.g.*, bacteria, animal, plant, or yeast. The vectors may optionally encode a "tag" sequence or sequences to facilitate protein purification. Convenient restriction enzyme cloning sites and suitable selective marker(s) are also optionally included. Such selective markers can be, for example, antibiotic resistance markers or markers which supply an essential nutritive growth factor to an otherwise deficient mutant host, *e.g.*, tryptophan, histidine, or leucine in the Yeast Two-Hybrid systems described below.

The term "recombinant vector" relates to a single- or double-stranded circular nucleic acid molecule that can be transfected into cells and replicated within or independently of a cell genome. A circular double-stranded nucleic acid molecule can be cut and thereby linearized upon treatment with appropriate restriction enzymes. An assortment of nucleic acid vectors, restriction enzymes, and the knowledge of the nucleotide sequences cut by restriction enzymes are readily available to those skilled in the art. A nucleic acid molecule encoding a desired product can be inserted into a vector by cutting the vector with restriction enzymes and ligating the two pieces together. Preferably the vector is an expression vector, *e.g.*, a shuttle expression vector as described above.

By "recombinant cell" is meant a cell possessing introduced or engineered nucleic acid sequences, *e.g.*, as described above. The sequence may be in the form of or part of a vector or may be integrated into the host cell genome. Preferably the cell is a bacterial cell.

In another aspect, the invention also provides methods for identifying and/or screening compounds "active on" at least one bacterial target of a bacteriophage inhibitor protein or RNA. Preferred embodiments involve contacting such a bacterial target or targets (*e.g.*, bacterial target proteins) with a test compound, and determining whether the compound binds to or reduces the level of activity of the bacterial target (*e.g.*, a bacterial target protein). Preferably this is done either *in vivo* (*i.e.*, in a cell-based assay) or *in vitro*, *e.g.*, in a cell-free system under approximately physiological conditions.

The compounds that can be used may be large or small, synthetic or natural, organic or inorganic, proteinaceous or non-proteinaceous. In preferred embodiments, the compound is a peptidomimetic, as described herein, a bacteriophage inhibitor protein or fragment or derivative thereof, preferably an "active portion", or a small molecule.

In preferred embodiments, the bacterial target is a target of a phage ORF identified herein, *e.g.*, *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

In particular embodiments, the methods include the identification of bacterial targets or the site of action of an inhibitor on a bacterial target as described above or otherwise described herein.

In embodiments involving binding assays, preferably binding is to a fragment or portion of a bacterial target protein, where the fragment includes less than 90%, 80%, 70%, 60%, 50%, 40%, or 30% of an intact bacterial target protein. Preferably,

the at least one bacterial target includes a plurality of different targets of bacteriophage inhibitor proteins, preferably a plurality of different targets. The plurality of targets can be in or from a plurality of different bacteria, but preferably is from a single bacterial species.

5 A "method of screening" refers to a method for evaluating a relevant activity or property of a large plurality of compounds (e.g., a bacteria-inhibiting activity), rather than just one or a few compounds. For example, a method of screening can be used to conveniently test at least 100, more preferably at least 1000, still more preferably at least 10,000, and most preferably at least 100,000 different compounds,
10 or even more.

In the context of this invention, the term "small molecule" refers to compounds having molecular mass of less than 2000 Daltons, preferably less than 1500, still more preferably less than 1000, and most preferably less than 600 Daltons. Preferably but not necessarily, a small molecule is not an oligopeptide.

15 In a related aspect or in preferred embodiments, the invention provides a method of screening for potential antibacterial agents by determining whether any of a plurality of compounds, preferably a plurality of small molecules, is active on at least one target of a bacteriophage inhibitor protein or RNA. Preferred embodiments include those described for the above aspect, including embodiments which involve
20 determining whether one or more test compounds bind to or reduce the level of activity of a bacterial target, and embodiments which utilize a plurality of different targets as described above.

The identification of bacteria-inhibiting phage ORFs and their encoded products also provides a method for identifying an active portion of such an encoded
25 product. This also provides a method for identifying a potential antibacterial agent by identifying such an active portion of a phage ORF or ORF product. In preferred embodiments, the identification of an active portion involves one or more of mutational analysis, deletion analysis, or analysis of fragments of such products. The method can also include determination of a 3-dimensional structure of an active
30 portion, such as by analysis of crystal diffraction patterns. In further embodiments, the method involves constructing or synthesizing a peptidomimetic compound, where the structure of the peptidomimetic compound corresponds to the structure of the active portion. In this context, "corresponds" means that the peptidomimetic compound structure has sufficient similarities to the structure of the active portion that
35 the peptidomimetic will interact with the same molecule as the phage protein and preferably will elicit at least one cellular response in common which relates to the inhibition of the cell by the phage protein.

In preferred embodiments, the ORF or ORF product is or is derived or obtained from *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014 or product thereof.

5 The methods for identifying or screening for compounds or agents active on a bacterial target of a phage-encoded inhibitor can also involve identification of a phage-specific site of action on the target.

Preferably in the methods for identifying or screening for compounds active on such a bacterial target, the target is uncharacterized; the target is from an uncharacterized bacterium from Table 1; the site of action is a phage-specific site of action.

Further embodiments include the identification of inhibitor phage ORFs and bacterial targets as in aspects above.

15 An "active portion" as used herein denotes an epitope, a catalytic or regulatory domain, or a fragment of a bacteriophage inhibitor protein that is responsible for, or a significant factor in, bacterial target inhibition. The active portion preferably may be removed from its contiguous sequences and, in isolation, still effect inhibition.

By "mimetic" is meant a compound structurally and functionally related to a reference compound that can be natural, synthetic, or chimeric. In terms of the present invention, a "peptidomimetic," for example, is a compound that mimics the activity-related aspects of the 3-dimensional structure of a peptide or polypeptide in a non-peptide compound, for example mimics the structure of a peptide or active portion of a phage- or bacterial ORF-encoded polypeptide.

A related aspect provides a method for inhibiting a bacterial cell by contacting 25 the bacterial cell with a compound active on a bacterial target of a bacteriophage inhibitor protein or RNA, where the target was uncharacterized. In preferred embodiments, the compound is such a protein, or a fragment or derivative thereof; a structural mimetic, *e.g.*, a peptidomimetic, of such a protein or fragment; a small molecule; the contacting is performed *in vitro*, the contacting is performed *in vivo* in an infected or at risk organism, *e.g.*, an animal such as a mammal or bird, for example, a human, or other mammal described herein; the bacterium is selected from a genus and/or species listed in Table 1; the bacteriophage inhibitor protein is uncharacterized; the bacteriophage inhibitor protein is from an uncharacterized phage listed in Table 1; the phage inhibitor protein is from one of *S. aureus* phage 44AHJD ORF 1, 9, or 12, 30 *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

In the context of targets in this invention, the term "uncharacterized" means that the target was not recognized as an appropriate target for an antibacterial agent prior to the filing of the present application or alternatively prior to the present invention. Such lack of recognition can include, for example, situations where the target and/or a nucleotide sequence encoding the target were unknown, situations where the target was known, but where it had not been identified as an appropriate target or as an essential cellular component, and situations where the target was known as essential but had not been recognized as an appropriate target due to a belief that the target would be inaccessible or otherwise that contacting the cell with a compound active on the target *in vitro* would be ineffective in cellular inhibition, or ineffective in treatment of an infection. Methods described herein utilizing bacterial targets, *e.g.*, for inhibiting bacteria or treating bacterial infections, can also utilize "uncharacterized target sites", meaning that the target has been previously recognized as an appropriate target for an antibacterial agent, but where an agent or inhibitor of the invention is used which acts at a different site than that at which the previously utilized antibacterial agent, *i.e.*, a phage-specific site. Preferably the phage-specific site has different functional characteristics from the previously utilized site. In the context of targets or target sites, the term "phage-specific" indicates that the target or site is utilized by at least one bacteriophage as an inhibitory target and is different from previously identified targets or target sites.

In the context of this invention, the term "bacteriophage inhibitor protein" refers to a protein encoded by a bacteriophage nucleic acid sequence which inhibits bacterial function in a host bacterium. Thus, it is a bacteria-inhibiting phage product.

In the context of this invention, the phrase "contacting the bacterial cell with a compound active on a bacterial target of a bacteriophage inhibitor protein" or equivalent phrases refer to contacting with an isolated, purified, or enriched compound or a composition including such a compound, but specifically does not rely on contacting the bacterial cell with an intact phage which encodes the compound. Preferably no intact phage are involved in the contacting.

Related aspects provide methods for prophylactic or therapeutic treatment of a bacterial infection by administering to an infected, challenged or at risk organism a therapeutically or prophylactically effective amount of a compound active on a target of a bacteriophage inhibitor protein or RNA, or as described for the previous aspect. Preferably the bacterium involved in the infection or risk of infection produces the identified target of the bacteriophage inhibitor protein or alternatively produces a homologous target compound. In preferred embodiments, the host organism is a plant or animal, preferably a mammal or bird, and more preferably, a human or other

mammal described herein. Preferred embodiments include, without limitation, those as described for the preceding aspect.

Compounds useful for the methods of inhibiting, methods of treating, and pharmaceutical compositions can include novel compounds, but can also include
5 compounds which had previously been identified for a purpose other than inhibition of bacteria. Such compounds can be utilized as described and can be included in pharmaceutical compositions.

In preferred embodiments of this and other aspects of the invention utilizing bacterial target sequences of a bacteriophage inhibitory ORF product, the target
10 sequence is encoded by a *Staphylococcus* nucleic acid coding sequence, preferably *S. aureus*, a *Streptococcus* nucleic acid coding sequence, preferably *Streptococcus pneumoniae*, or *Enterococcus* nucleic acid coding sequence. Possible target sequences are described herein by reference to sequence source sites.

The amino acid sequence of a polypeptide target is readily provided by
15 translating the corresponding coding region. For the sake of brevity, the sequences are not reproduced herein. For the sake of brevity, the sequences are described by reference to the GenBank entries instead of being written out in full herein. In cases where the TIGR or GenBank entry for a coding region is not complete, the complete sequence can be readily obtained by routine methods, e.g., by isolating a clone in a
20 phage host genomic library, and sequencing the clone insert to provide the relevant coding region. The boundaries of the coding region can be identified by conventional sequence analysis and/or by expression in a bacterium in which the endogenous copy of the coding region has been inactivated and using subcloning to identify the functional start and stop codons for the coding region.

25 In the context of nucleic acid or amino acid sequences of this invention, the term "corresponding" indicates that the sequence is at least 95% identical, preferably at least 97% identical, and more preferably at least 99% identical to a sequence from the specified phage genome, a ribonucleotide equivalent, a degenerate equivalent (utilizing one or more degenerate codons), or a homologous sequence, where the
30 homolog provides functionally equivalent biological function.

By "treatment" or "treating" is meant administering a compound or pharmaceutical composition for prophylactic and/or therapeutic purposes. The term "prophylactic treatment" refers to treating a patient or animal that is not yet infected but is susceptible to or otherwise at risk of a bacterial infection. The term "therapeutic
35 treatment" refers to administering treatment to a patient already suffering from infection.

The term "bacterial infection" refers to the invasion of the host organism, animal or plant, by pathogenic bacteria. This includes the excessive growth of bacteria which are normally present in or on the body of the organism, but more generally, a bacterial infection can be any situation in which the presence of a bacterial population(s) is damaging to a host organism. Thus, for example, an organism suffers from a bacterial population when excessive numbers of a bacterial population are present in or on the organism's body, or when the effects of the presence of a bacterial population(s) is damaging to the cells, tissue, or organs of the organism.

The terms "administer", "administering", and "administration" refer to a method of giving a dosage of a compound or composition, *e.g.*, an antibacterial pharmaceutical composition, to an organism. Where the organism is a mammal, the method is, *e.g.*, topical, oral, intravenous, transdermal, intraperitoneal, intramuscular, or intrathecal. The preferred method of administration can vary depending on various factors, *e.g.*, the components of the pharmaceutical composition, the site of the potential or actual bacterial infection, the bacterium involved, and the infection severity.

The term "mammal" has its usual biological meaning referring to any organism of the Class Mammalia of higher vertebrates that nourish their young with milk secreted by mammary glands, *e.g.*, mouse, rat, and, in particular, human, bovine, sheep, swine, dog, and cat.

In the context of treating a bacterial infection a "therapeutically effective amount" or "pharmaceutically effective amount" indicates an amount of an antibacterial agent, *e.g.*, as disclosed for this invention, which has a therapeutic effect. This generally refers to the inhibition, to some extent, of the normal cellular functioning of bacterial cells that renders or contributes to bacterial infection.

The dose of antibacterial agent that is useful as a treatment is a "therapeutically effective amount." Thus, as used herein, a therapeutically effective amount means an amount of an antibacterial agent that produces the desired therapeutic effect as judged by clinical trial results and/or animal models. This amount can be routinely determined by one skilled in the art and will vary depending on several factors, such as the particular bacterial strain involved and the particular antibacterial agent used.

In connection with claims to methods of inhibiting bacteria and therapeutic or prophylactic treatments, "a compound active on a target of a bacteriophage inhibitor protein" or terms of equivalent meaning differ from administration of or contact with an intact phage naturally encoding the full-length inhibitor compound. While an intact phage may conceivably be incorporated in the present methods, the method at

least includes the use of an active compound as specified different from a full length inhibitor protein naturally encoded by a bacteriophage and/or a delivery or contacting method different from administration of or contact with an intact phage encoding the full-length protein. Similarly, pharmaceutical compositions described herein at least
5 include an active compound different from a full-length inhibitor protein naturally encoded by a bacteriophage or such a full-length protein is provided in the composition in a form different from being encoded by an intact phage. Preferably the methods and compositions do not include an intact phage.

In accord with the above aspects, the invention also provides antibacterial
10 agents and compounds active on bacterial targets of bacteriophage inhibitor proteins or RNAs, where the target was uncharacterized as indicated above. As previously indicated, such active compounds include both novel compounds and compounds which had previously been identified for a purpose other than inhibition of bacteria. Such previously identified biologically active compounds can be used in
15 embodiments of the above methods of inhibiting and treating. In preferred embodiments, the targets, bacteriophage, and active compound are as described herein for methods of inhibiting and methods of treating. Preferably the agent or compound is formulated in a pharmaceutical composition which includes a pharmaceutically acceptable carrier, excipient, or diluent. In addition, the invention provides agents,
20 compounds, and pharmaceutical compositions where an active compound is active on an uncharacterized phage-specific site.

In preferred embodiments, the target is as described for embodiments of aspects above.

Likewise, the invention provides a method of making an antibacterial agent.
25 The method involves identifying a target of a bacteriophage inhibitor polypeptide or protein or RNA, screening a plurality of compounds to identify a compound active on the target, and synthesizing the compound in an amount sufficient to provide a therapeutic effect when administered to an organism infected by a bacterium naturally producing the target. In preferred embodiments, the identification of the target and
30 identification of active compounds include steps or methods and/or components as described above (or otherwise herein) for such identification. Likewise, the active compound can be as described above, including fragments and derivatives of phage inhibitor proteins, peptidomimetics, and small molecules. As recognized by those skilled in the art, peptides can be synthesized by expression systems and purified, or
35 can be synthesized artificially. In preferred embodiments the inhibitory phage ORF products is from *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus*

pneumoniae phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

As indicated above, sequence analysis of nucleotide and/or amino acid sequences can beneficially utilize computer analysis. Thus, in additional aspects the invention provides computer-related hardware and media and methods utilizing and incorporating sequence data from uncharacterized phage, *e.g.*, uncharacterized phage listed in Table 1, preferably at least one of *Staphylococcus aureus* phage *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014, or 44 AHJD, *Enterococcus* sp. phage 182, or *Streptococcus pneumoniae* phage Dp-1. In general, such aspects can facilitate the above-described aspects. Various embodiments involve the analysis of genetic sequence and encoded products, as applied to the evaluating bacteriophage inhibitor ORFs and compounds and fragments related thereto. The various sequence analyses, as well as function analyses, can be used separately or in combination, as well as in preceding aspects and embodiments. Use in combination is often advantageous as the additional information allows more efficient prioritizing of phage ORFs for identification of those ORFs that provide bacteria-inhibiting function.

In one aspect, the invention provides a computer-readable device which includes at least one recorded amino acid or nucleotide sequence corresponding to one of the specified phage and a sequence analysis program for analyzing a nucleotide and/or amino acid sequence. The device is arranged such that the sequence information can be retrieved and analyzed using the analysis program. The analysis can identify, for example, homologous sequences or the indicated %s of the phage genome and structural motifs. Preferably the sequence includes at least 1 phage ORF or encoded product, more preferably at least 10%, 20%, 30%, 40%, 50%, 70%, 90%, or 100% of the genomic phage ORFs and/or equivalent cDNA, RNA, or amino acid sequences. Preferably the sequence or sequences in the device are recorded in a medium such as a floppy disk, a computer hard drive, an optical disk, computer random access memory (RAM), or magnetic tape. The program may also be recorded in such medium. The sequences can also include sequences from a plurality of different phage.

In this context, the term "corresponding" indicates that the sequence is at least 95% identical, preferably at least 97% identical, and more preferably at least 99% identical to a sequence from the specified phage genome, a ribonucleotide equivalent, a degenerate equivalent (utilizing one or more degenerate codons), or a homologous sequence, where the homolog provides functionally equivalent biological function.

Similarly, the invention provides a computer analysis system for identifying biologically important portions of a bacteriophage genome. The system includes a data storage medium, *e.g.*, as identified above, which has recorded thereon a nucleotide sequence corresponding to at least a portion of at least one uncharacterized bacteriophage genome, a set of program instructions to allow searching of the sequence or sequences to analyze the sequence, and an output device where the portion includes at least the sequence length as specified in the preceding aspect. The output device is preferably a printer, a video display, or a recording medium. More than one output device may be included. For each of the present computer-related aspects, the bacteriophage are preferably selected from the uncharacterized phage listed in Table 1, more preferably from bacteriophage 77, 3A, 96, 44 AHJD (*S. aureus*), Dp-1 (*Streptococcus pneumoniae*), or 182 (*Enterococcus*).

In keeping with the computer device aspects, the invention also provides a method for identifying or characterizing a bacteriophage ORF by providing a computer-based system for analyzing nucleotide or amino acid sequences, *e.g.*, as describe above. The system includes a data storage medium which has recorded a sequences or sequences as described for the above devices, a set of instructions as in the preceding aspect, and an output device as in the preceding aspect. The method further involves analyzing at least one sequence, and outputting the analysis results to at least one output device.

In preferred embodiments, the analysis identifies a sequence similarity or homology with a sequence or sequences selected from bacterial ORFs encoding products with related biological function; ORFs encoding known inhibitors; and essential bacterial ORFs. Preferably the analysis identifies a probable biological function based on identification of structural elements or characteristic or signature motifs of an encoded product or on sequence similarity or homology. Preferably the uncharacterized bacteriophage is from Table 1, more preferably at least one of bacteriophage 77, 3A, 96, 44 AHJD (*S. aureus*), Dp-1 (*Streptococcus pneumoniae*), or 182 (*Enterococcus*). In preferred embodiments, the method also involves determining at least a portion of the nucleotide sequence of at least one uncharacterized bacteriophage as indicated, and recording that sequence on data storage medium of the computer-based system. In preferred embodiments, the analysis identifies a sequence similarity of homology with a *S. aureus* phage 44AHJD ORF 1, 9, or 12, *Streptococcus pneumoniae* phage Dp-1 ORF 001, 002, 004, 008, 010, 013, 016, 021, 029, 030, 038, or 041, or *Enterococcus* sp. phage 182 ORF 002, 008, or 014.

As used in the claims to describe the various inventive aspects and embodiments, "comprising" means including, but not limited to, whatever follows the word "comprising". Thus, use of the term "comprising" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present. By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of". Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements may be present. By "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present depending upon whether or not they affect the activity or action of the listed elements.

Further embodiments will be apparent from the following Detailed Description and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1A and 1B are flow schematics showing the manipulations used to convert pT0021, an arsenite inducible vector containing the luciferase gene, into pTHA or pTM, two *ars* inducible vectors. Vector pTHA contains BamHI, SalI, and HindIII cloning sites and a downstream HA epitope tag. Vector pTM contains BamHI and HindIII cloning sites and no HA epitope tag.

FIGURE 2 is a schematic representation of the cloning steps involved to place the DNA segments of any of ORFs 17/ 19/ 43/ 102/104/182 or other sequences into pTHA to assess inhibitory potential. For subcloning into pTM or pT0021, individual ORFs were amplified by the PCR using oligonucleotides targeting the ATG and stop codons of the ORFs. Using this strategy, BamHI and HindIII sites were positioned immediately upstream or downstream, respectively of the start and stop codons of each ORF. Following digestion with BamHI and HindIII, the PCR fragments were subcloned into the same sites of pT0021 or pTM. Clones were verified by PCR and direct sequencing.

FIGURE 3 shows a schematic representation of the functional assays used to characterize the bactericidal and bacteriostatic potential of all predicted ORFs (>33 amino acids) encoded by bacteriophage 77. Fig. 3A) Functional assay on semi-solid support media. Fig. 3B) Functional assay in liquid culture.

FIGURE 4A, B, and C is a bar graph showing the results of a screen in liquid media to assess bacteriostatic or bactericidal activity of 93 predicted ORFs (>33 amino acids) encoded by bacteriophage 77. Growth inhibition assays were performed as detailed in the Detailed Description. The relative growth of *Staphylococcus aureus* transformants harboring a given bacteriophage 77 ORF (identified on the bottom of the graph), in the absence or presence of arsenite, is plotted relative to growth of a *Staphylococcus aureus* transformant containing ORF 5, a non-toxic bacteriophage 77 ORF (which is set at 100%). Each bar represents the average obtained from three *Staph A* transformants grown in duplicate. Bacteriophage 77 ORFs showing significant growth inhibition consist of ORFs 17, 19, 102, 104, and 182.

FIGURE 5 shows a block diagram of major components of a general purpose computer.

FIGURE 6 shows an ORF map for *Streptococcus pneumoniae* bacteriophage Dp-1 showing the ORF identifiers, genomic locations, and orientations of the 85 identified ORFs that were found to have ribosomal binding sites and thus are expected to be expressed.

FIGURE 7 shows a schematic representation of the arsenite-inducible expression system present in a shuttle vector designed to express individual *Streptococcus* bacteriophage Dp-1 ORFs in *Streptococcus*. Various modifications can be readily made to such a vector, or other vectors can be readily constructed to provide inducible expression of ORFs in a particular host bacterium using well-known techniques.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The invention may be more clearly understood from the following description.

5 The tables will first be briefly described.

Table 1 is a listing of a large number of available bacteriophage that can be readily obtained and used in the present invention.

Table 2 shows the complete nucleotide sequence of the genome of *Staphylococcus aureus* bacteriophage 77.

10 Table 3 shows a list of all the ORFs from Bacteriophage 77 that were screened in the functional assay to identify those with anti-microbial activity.

Table 4 shows the predicted nucleotide sequence, predicted amino acid sequence, and physiochemical parameters of ORF 17/ 19/ 43/ 102/ 104/ 182]. These include the primary amino acid sequence of the predicted protein, the average
15 molecular weight, amino acid composition, theoretical pI, hydrophobicity map, and predicted secondary structure map.

Table 5 shows homology search results. BLAST analysis was performed with ORFs 17/ 19/ 43/ 102/ 104/ 182 against NCBI non-redundant nucleotide and Swissprot databases. The results of this search indicate that: I) ORF 17 has no
20 significant homology to any gene in the NCBI non-NCBI non-redundant nucleotide database, II) ORF 19 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 59 of bacteriophage phi PVL, III) ORF 43 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 39 of phi PVL, IV) ORF 102 has
25 significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 38 of phi PVL, V) ORF 104 has no significant homology to any gene in the NCBI non-redundant nucleotide database, VI) ORF 182 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 39 of phi PVL.

30 Table 6 is a table from Alberts et al., MOLECULAR BIOLOGY OF THE CELL 3rd ed., showing the redundancy of the "universal" genetic code.

Table 7 shows the complete nucleotide sequence of *Staphylococcus aureus* bacteriophage 3A.

Table 8 is a listing of the ORFs identified in *Staphylococcus aureus* bacteriophage 3A.

Table 9 shows the complete nucleotide sequence of *Staphylococcus aureus* bacteriophage 96.

5 Table 10 is a listing of the ORFs identified in *Staphylococcus aureus* bacteriophage 96.

Table 11 is a listing of sequences deposited in the NCBI public database (GeneBank) for bacteriophage listed in Table 1.

10 Table 12 is a listing of phage which encode a known lysis function , including the identified lysis gene.

Table 13 is a listing of bacteriophage which encode holin genes, where holin genes encode proteins which form pores and eventually enable other enzymes to kill the host bacterium.

Table 14 is a listing of bacteriophage which encode kil genes.

15 Table 15 is a list of *Staphylococcus aureus* sequences identified by accession number which may include sequences from genes coding for target sequences for the phage 77-encoded antimicrobial proteins or peptides. The sequences were obtained by searching GenBank for listings.

20 Table 16 shows the nucleotide sequence of the genome of *Staphylococcus aureus* phage 44 AHJD.

Table 17 lists and shows the sequence position of the 73 ORFs predicted to be encoded by *Staphylococcus aureus* bacteriophage 44 AHJD that are greater than 33 amino acids.

25 Table 18 shows the ORF sequences and putative amino acid sequences for the *Staphylococcus aureus* bacteriophage 44AHJD ORFs greater than 33 amino acids.

Table 19 shows the similarities in sequence identified between predicted *Staphylococcus aureus* bacteriophage 44 AHJD ORFs and sequences present in public databases.

30 Table 20 shows the homology alignments between predicted *Staphylococcus aureus* bacteriophage 44AHJD ORFs and the corresponding protein sequences present in public sequence databases.

Table 21 shows the complete nucleotide sequence of the genome of *Enterococcus* bacteriophage 182.

35 Table 22 lists and shows the sequence position of the 80 ORFs identified in bacteriophage 182 and that are greater than 33 amino acids.

Table 23 shows the nucleotide and predicted amino acid sequence of all 80 ORFs identified in bacteriophage 182.

Table 24 shows the similarities identified to date in sequence between *Enterococcus* phage 182 ORFs greater than 33 amino acids and sequences present in public sequence databases.

Table 25 shows the predicted amino acid sequence as well as the predicted secondary structures map for two *Enterococcus* bacteriophage 182 ORFs.

Table 26 shows the homology alignments between predicted *Enterococcus* bacteriophage 182 ORFs and the corresponding protein sequences present in public sequence databases.

Table 27 list *Enterococcus* sequences listed in GenBank providing possible Enterococcal target sequences for inhibitory *Enterococcus* bacteriophage 182 ORFs and other compounds with antibacterial activity.

Table 28 shows the complete nucleotide sequence of the genome of *Streptococcus* bacteriophage Dp-1.

Table 29 lists and shows sequence position of the 273 ORFs identified in Pneumococcal bacteriophage Dp-1 that are greater than 33 amino acids, 85 of which are predicted to be expressed in Dp-1 as having a ribosomal binding site. That set of 85 ORFs is shown in the attached drawings.

Table 30 shows the nucleotide and predicted amino acid sequence of all 273 ORFs identified in bacteriophage Dp-1 that are identified as being expressed.

Table 31 shows the similarities identified in sequence between *Streptococcus* phage Dp-1 ORFs greater than 33 amino acids and sequences present in public sequence databases.

Table 32 shows the 4731 bp sequence of Dp-1 published by Sheehan et al., 1997).

Table 33 lists *Streptococcus pneumoniae* sequences listed in GenBank providing possible target sequences for inhibitory *Streptococcus pneumoniae* bacteriophage Dp-1 ORFs and other compounds with antibacterial activity

Background:

As indicated above, the present invention is concerned, in part, with the use of bacteriophage coding sequences and the encoded polypeptides or RNA transcripts to identify bacterial targets for potential new antibacterial agents. Thus, the invention concerns the selection of relevant bacteria. Particularly relevant bacteria are those which are pathogens of a complex organism such as an animal, e.g., mammals,

reptiles, and birds, and plants. Examples include *Staphylococcus aureus*, *Enterococcus* species, and *Streptococcus pneumoniae*. However, the invention can be applied to any bacterium (whether pathogenic or not) for which bacteriophage are available or which are found to have cellular components closely homologous to components targeted by phage of another bacterium.

Thus, the invention also concerns the bacteriophage which can infect a selected bacterium. Identification of ORFs or products from the phage which inhibit the host bacterium both provides an inhibitor compound and allows identification of the bacterial target affected by the phage-encoded inhibitor. Such targets are thus identified as potential targets for development of other antibacterial agents or inhibitors and the use of those targets to inhibit those bacteria. As indicated above, even if such a target is not initially identified in a particular bacterium, such a target can still be identified if a homologous target is identified in another bacterium. Usually, but not necessarily, such another bacterium would be a genetically closely related bacterium. Indeed, in some cases, a phage-encoded inhibitor can also inhibit such a homologous bacterial cellular component.

The demonstration that bacteriophage have adapted to inhibiting a host bacterium by acting on a particular cellular component or target provides a strong indication that that component is an appropriate target for developing and using antibacterial agents, e.g., in therapeutic treatments. Thus, the present invention provides additional guidance over mere identification of bacterial essential genes, as the present invention also provides an indication of accessibility of the target to an inhibitor, and an indication that the target is sufficiently stable over time (e.g., not subject to high rates of mutation) as phage acting on that target were able to develop and persist. Thus, the present invention identifies a subset of essential cellular components which are particularly likely to be appropriate targets for development of antibacterial agents.

The invention also, therefore, concerns the development or identification of inhibitors of bacteria, in addition to the phage-encoded inhibitory proteins (or RNA transcripts), which are active on the targets of bacteriophage-encoded inhibitors. As described herein, such inhibitors can be of a variety of different types, but are preferably small molecules.

The following description provides preferred methods for use in the various aspects of the invention. However, as those skilled in the art will readily recognize, other approaches can be used to obtain and process relevant information. Thus the invention is not limited to the specifically described methods. In addition, the following description provides a set of steps in a particular order. That series of steps

describes the overall development involved in the present invention. However, it is clear that individual steps or portions of steps may be usefully practiced separately, and, further, that certain steps may be performed in a different order or even bypassed if appropriate information is already available or is provided by other sources or methods.

Selecting and Growing Phage, and Isolating DNA

Conceptually, the first step involves selecting bacterial hosts of interest. Preferably, but not necessarily, such hosts will be pathogens of clinical importance. Alternatively, because bacteria all share certain fundamental metabolic and structural features, these features can be targeted for study in one strain, for example a nonpathogenic one, and extrapolated to similarly succeed in pathogenic ones. Nonpathogenic strains may also exhibit initial advantages in being not only less dangerous, but also, for example, in having better growth and culturing characteristics and/or better developed molecular biology techniques and reagents. Consequently, advantageously the invention provides the ability target virtually any bacteria, but preferably pathogenic bacteria, with antimicrobial compounds designed and/or developed using bacteriophage inhibitory proteins and peptides from phage with non-pathogenic and/or pathogenic hosts.

We have selected *Staphylococcus aureus*, *Streptococcus pneumoniae*, various *Enterococci*, and *Pseudomonas aeruginosa* as initial exemplary pathogens. These bacteria are a major cause of morbidity and mortality in hospital-based infections, and the appearance of antibiotics resistance in all three organisms makes it increasingly difficult to treat benign infections involving these organisms. Such infections can include, for example, otitis media, sinusitis, and skin, and airway infections (Neu, H.C. (1992). *Science* 257, 1064-1073). However, the approach described below is clearly applicable to any human bacterial pathogens including but not restricted to *Mycobacterium tuberculosis*, *Nisseria gonorrhoeae*, *Haemophilus influenza*, *Acinobacter*, *Escherichia coli*, *Shigella dysenteria*, *Streptococcus pyogenes*, *Helicobacter pylori*, and *Mycoplasma* species. This invention can also be applied to the discovery of anti-bacterial compounds directed against pathogens of animals other than humans, for example, sheep, cattle, swine, dogs, cats, birds, and reptiles. Similarly, the invention is not limited to animals, but also applies to plants and plant pathogens.

In general, the bacteria are grown according to standard methodologies employed in the art, including solid, semi-solid or liquid culturing, which procedures can be found in or extrapolated from standard sources such as Maloy, S.R., Stewart,

- V.J., and Taylor, R.K. Genetic Analysis of Pathogenic Bacteria (1996) Cold Spring Harbor Laboratory Press, or Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; or Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J. Culture conditions are selected which are adapted to the particular bacterium generally using culture conditions known in the art as appropriate, or adaptations of those conditions.

- Nucleic acids within these bacteria can be routinely extracted through common procedures such as described in the above-referenced manuals and as generally known to those skilled in the art. Those nucleic acid stocks can then be used to practice the other inventive aspects described below.

Selection and Growth of Bacteriophage, and Isolation of DNA

- The second step involves assembling a group of bacteriophages (phage collection) for one or more of the targeted bacterial hosts. While the invention can be utilized with a single bacteriophage for a pathogen or other bacterium, it is preferable to utilize a plurality of phage for each bacterium, as comparisons between a plurality of such phage provides useful additional information. Non-limiting examples of phage and sources for some of the above-mentioned pathogenic bacteria are found in Table 1. The criteria used to select such phages is that they are infectious for the microbe targeted, and replicate in, lyse, or otherwise inhibit growth of the bacterium in a measurable fashion. These phages can be very different from one another (representing different families), as judged by criteria such as morphology (head, tail, plate, etc.), and similarity of genome nucleotide sequence (cross-hybridization). Since such diverse bacteriophages are expected to block bacterial host metabolism and ultimately inhibit by a variety of mechanisms, their combined study will lead to the identification of different mechanisms by which the phages independently inhibit bacterial targets. Examples include degradation of host DNA (Parson K.A., and Snustad, D.P. (1975). *J. Virol.* 15, 221-444) and inhibition of host RNA transcription (Severinova, E., Severinov, K. and Darst, S.A. (1998). *J.Mol. Biol.* 279, 9-18). This, in turn, yields novel information on phage proteins that can inhibit the targeted microbe. As explained below, this 1) forms the basis of novel drug discovery efforts based on knowledge of the primary amino acid sequence of the phage inhibitor protein (e.g., peptide fragments or peptidomimetics) and/or 2) leads to the identification of bacterial biochemical pathways, the proteins of which are essential or significant for survival of the targeted microbe, and which enzymatic steps or

chemical reactions can be targeted by classical drug discovery methods using molecular inhibitors, for example, small molecule inhibitors.

Bacteriophage are generally either of two types, lytic or filamentous, meaning they either outright destroy their host and seek out new hosts after replication, or else continuously propagate and extrude progeny phage from the same host without destroying it. Regardless of the phage life cycle and type, preferred embodiments incorporate phage which impede cell growth in measurable fashion and preferably stop cell growth. To this end, lytic phage are preferred, although certain nonlytic species may also suffice, *e.g.*, if sufficiently bacteriostatic.

Various procedures that are commonly understood by those of skill in the art can be routinely employed to grow, isolate, and purify phage. Such procedures are exemplified by those found in such common laboratory aids such as Maloy, S.R., Stewart, V.J., and Taylor, R.K. Genetic Analysis of Pathogenic Bacteria (1996) Cold Spring Harbor Laboratory Press; Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; and Ausubel, F.M. et al. (eds.) (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J. The techniques generally involve the culturing of infected bacterial cells that are lysed naturally and/or chemically assisted, for example, by the use of an organic solvent such as chloroform that destroys the host cells thereby liberating the phage within. Following this, the cellular debris is centrifuged away from the supernatant containing the phage particles, and the phage then subsequently and selectively precipitated out of the supernatant using various methods usually employing the use of alcohols and/or other chemical compounds such as polyethylene glycol (PEG). The resulting phage can be further purified using various density gradient/centrifugation methodologies. The resulting phage are then chemically lysed, thereby releasing their nucleic acids that can be conveniently precipitated out of the supernatant to yield a viral nucleic acid supply of the phage of interest.

Exemplary bacteriophage are indicated in Table 1, along with sources where those phage may be obtained.

Exemplary bacteria include the reference bacteria for the identified bacteriophage, available from the same sources.

Characterizing Bacteriophage Genomes for ORFs

The third step involves systematically characterizing the genetic information contained in the phage genome. Within this genetic information is the sequence of all RNAs and proteins encoded by the phage, including those that are essential or

instrumental in inhibiting their host. This characterization is preferably done in a systematic fashion. For example, this can be done by first isolating high molecular weight genomic DNA from the phage using standard bacterial lysis methods, followed by phage purification using density gradient ultracentrifugation, and extraction of nucleic acid from the purified phage preparation. The high molecular weight DNA is then analyzed to determine its size and to evaluate a proper strategy for its sequencing. The DNA is broken down into smaller size fragments by sonication or partial digestion with frequently cutting restriction enzymes such as Sau3A to yield predominantly 1 to 2 kilobase length DNA, which DNA can then be resolved by gel electrophoresis followed by extraction from the gel.

The ends of the fragments are enzymatically treated to render them suitable for cloning and the pools of fragments are cloned in a bacterial plasmid to generate a library of the phage genome. Several hundred of these random DNA fragments contained in the plasmid vector are isolated as clones after introduction into an appropriate bacterium, usually *Escherichia coli*. They are then individually expanded in culture and the DNA from each individual clone is purified. The nucleotide sequences of the inserts of these clones are determined by standard automated or manual methods, using oligonucleotide primers located on either side of the cloning site to direct polymerase mediated sequencing (e.g., the Sanger sequencing method or a modification of that method). Other sequencing methods can also be used.

The sequence of individual clones is then deposited in a computer, and specific software programs (for example, Sequencher™, Gene Codes Corp.) are used to look for overlap between the various sequences, resulting in ordering of contig sequences and ultimately providing the complete sequence of the entire bacteriophage genome (one such example is given in Table 2 for *Staphylococcus aureus* bacteriophage 77; others are also provided herein). This complete nucleotide sequence is preferably determined with a redundancy of at least 3- to 5-fold (number of independent sequencing events covering the same region) in order to minimize sequencing errors.

Preferably, the bacterial strain used as a phage host should not possess any other innate plasmids, transposons, or other phage or incompatible sequences that would complicate or otherwise make the various manipulations and analyses more difficult.

Commercially available computer software programs are used to translate the nucleotide sequence of the phage to identify all protein sequences encoded by the phage (hereafter called open reading frames or ORFs). (Customized software can clearly also be used.) As phages are known to transcribe their genome into RNA from

both strands, in both directions, and sometimes in more than one frame for the same sequence, this exercise is done for both strands and in all six possible reading frames. As evolutionary constraints have forced the phage to conserve all of its vital protein sequences in as small a genome as possible, it is straightforward to identify all the proteins encoded by the phage by simple examination of the 6 translation frames of the genome. Once these ORFs are identified, they are cataloged into a phage proteome database (Table 3 lists ORFs identified from phage 77; ORF lists are also provided for other exemplary phage). This analysis is preferably performed for each phage under study. The process of ORF identification can be varied depending on the desired results. For example, the minimum length for the putative encoded polypeptide can be varied, and/or putative coding regions that have an associated Shine-Dalgarno sequence can be selected. In the case of phage 77 ORFs, such parameter adjustment was performed and resulted in the identification of ORFs as listed herein. Different parameters had resulted in the identification of the ORFs listed in the preceding U.S. Provisional Application 60/110,992, filed December 3, 1998, which is hereby incorporated by reference in its entirety.

Exemplary phage 77 ORFs identified in that provisional application and as identified herein are shown in the following table:

ORF ID from 60/110,992	Genomic position	a.a. size	Start codon	ORF ID from 241/190	Genomic position	a.a. size	Start codon
77ORF016	2369-24024	251	TTG	77ORF017	23269-23982	237	ATG
77ORF019	39845-40501	218	ATA	77ORF019	39851-40501	216	ATG
77ORF050	29268-29564	98	ATG	77ORF182	29268-29564	98	ATG
77ORF050	29268-29564	98	ATG	77ORF043	29304-29564	86	ATG
77ORF067	34312-34551	79	CTG	77ORF104	34393-34551	52	ATG
77ORF146	29051-29212	53	ATG	77ORF102	29051-29212	53	ATG

Identifying and Characterizing Inhibitory Phage ORFs

The fourth step entails identifying the phage protein or proteins or RNA transcripts that have the ability to inhibit their bacterial hosts. This can be accomplished, for example, by either or both of two non-mutually exclusive methods. The first method makes use of bioinformatics. Over the past few years, a large amount of nucleotide sequence information and corresponding translated products have become available through large genome sequencing projects for a variety of organisms including mammals, insects, plants, unicellular eukaryotes (yeast and fungi), as well as several bacterial genomes such as *E. coli*, *Mycobacterium tuberculosis*, *Bacillus subtilis*, *Staphylococcus aureus* and many others. Such sequences have been deposited in public databases (for example, non-redundant

sequence database at GenBank and SwissProt protein sequence database) (http://www.ncbi.nlm.nih.gov)) and can be freely accessed to compare any specific query sequence to those present in such databases. For example, GenBank contains over 1.6 billion nucleotides corresponding to 2.3 million sequence records. Several computer programs and servers (e.g., TBLASTN) have been created to allow the rapid identification of homology between any given sequence from one organism to that of another present in such databases, and such programs are public and available free of charge.

In addition, it has been well established that basic biochemical pathways can be conserved in very distant organisms (for example bacteria and man), and that the proteins performing the various enzymatic steps in these pathways are themselves conserved at the amino acid sequence level. Thus, proteins performing similar functions (e.g. DNA repair, RNA transcription, RNA translation) have frequently preserved key structural signatures, identifiable by similarities across regions of proteins (domains and motifs). The antimicrobials of the present invention will preferably target features and targets that are highly characteristic or conserved in microbes, and not higher organisms.

Most genomes encode individual proteins or groups of proteins that can be assembled into protein families that have been evolutionarily conserved. Therefore, similarity between a new query sequence and that of a member of a protein family (reference sequences from public databases) can immediately suggest a biochemical function for the novel query sequence, which in our case is a phage ORF.

The sequence homology between individual members of evolutionarily distant members of a protein family is usually not randomly distributed along the entire length of the sequence but is often clustered into "motifs" and "domains". These correspond to key three-dimensional folds that form key catalytic and/or regulatory structures that perform key biochemical function(s) for the group of proteins. Commercially available computer software programs can identify such motifs in a new query sequence, again providing functional information for the query sequence. Such structural and functional motifs have also been derived from the combined analysis of primary sequence databases (protein sequences) and protein structure databases (X-ray crystallography, nuclear magnetic resonance) using so-called "threading" methods (Rost B, I and Sander C. (1996) *Ann. Rev. Biophys. Biomol. Struct.* 25, 113-136).

Such motifs and folds are themselves deposited in public databases which can be directly accessed (for example, SwissProt database; 3D-ALI at EMBL, Heidelberg; PROSITE). This basic exercise leads to a structural homology map in which each of

the phage ORFs has been probed for such similarities, and where initial structural and functional hits are identified (selected examples of sequence homologies detected between individual ORFs from the genome of *Staphylococcus aureus* bacteriophage 77 and sequences deposited in public databases are shown in Table 5 for ORFs
 5 17/19/43/102/104/182).

This analysis can point out phage proteins with similarity to proteins from other phages (such as those for *E. coli*) playing an important role in the basic biochemical pathways of the phage (such as DNA replication, RNA transcription, tRNAs, coat protein and assembly). Selected examples of such proteins include
 10 integrase and capsid protein. Therefore, this analysis enables identification and elimination of non-essential ORFs as candidates for an inhibitor function, as well as the identification of (potentially) useful ones.

In addition, this analysis can point out specific ORFs as possible inhibitor ORFs. For example these ORFs may encode proteins or enzymes that alter bacterial
 15 cell structure, metabolism or physiology, and ultimately viability. Examples of such proteins present in the genome of *Staphylococcus aureus* bacteriophage 77 include orf14 (deoxyuridine triphosphatase from bacteriophage T5), and orf15 (sialidase). (These ORF identifications are as listed in provisional application 60/110,992.) Other examples include ORFs 9 and 12 of *S. aureus* phage 44 AHJD, which encode the
 20 putative lysis functions found in many bacteriophages – a “holin” and an “amidase”.

In addition, it is well known that bacterial and eukaryotic viruses can usurp pathways from their host in order to use them to their advantage in blocking host cellular pathways upon infection. The phage can achieve this by 1) directly producing an inhibitor of a key host pathway (e.g. T7 gene 0.5 and 2), 2) directly producing a
 25 novel activity (e.g. T4 DNA polymerase), and 3) altering concentrations of cell components by producing similar functions (e.g. T4 transfer RNAs). The identification of sequence similarity between phage ORFs and bacterial host genome sequences will be highly indicative of such a mechanism. (Selected examples of such homologies are listed in Figure 4 of the provisional application 60/110,992 and
 30 include orf4 (homologous to autolysin), orf20 (hypothetical protein from *Staphylococcus aureus*) and orf29 (hypothetical protein from *Staphylococcus aureus*.) These ORFs can be analyzed by a standard biochemical approach to directly test their inhibitor functions (e.g., as described below).

Alternatively, a homology search may reveal that a given phage ORF is related
 35 to a protein present in the databases having an activity known to be inhibitory, (e.g., inhibitor of host RNA polymerase by *E. coli* bacteriophage T7. Such a finding would implicate the phage ORF product in a related activity. This will also suggest that a

new antimicrobial could be derived by a mimetic approach (e.g., peptidomimetic) imitating this function or by a small molecule inhibitor to the bacterial target of the phage ORF, or any steps in the relevant host metabolic pathway, e.g., high throughput screening of small molecule libraries. Selected examples of such similarity between
5 ORFs of *Staphylococcus aureus* bacteriophage 77 and proteins with inhibitor functions for bacterial hosts are listed in Figure 4 of the provisional application 60/110,992. These include orf9 (similar to bacteriophage P1 *kilA* function), and orf4 (autolysin of *Staphylococcus aureus*, amidase enzymatic activity).

A reason for the biochemical study of individual ORFs for inhibitor function is
10 that their expression or overexpression will block cellular pathways of the host, ultimately leading to arrest and/or inhibition of host metabolism. In addition, such ORFs can alter host metabolism in different ways, including modification of pathogenicity. Therefore, individual ORFs identified above are expressed, preferably overexpressed, in the host and the effect of this expression or overexpression on host
15 metabolism and viability is measured. This approach can be systematically applied to every ORF of the phage, if necessary, and does not rely on the absolute identification of candidate ORFs by bioinformatics. Individual ORFs are resynthesized from the phage genomic DNA, e.g., by the polymerase chain reaction (PCR), preferably using oligonucleotide primers flanking the ORF on either side. These single ORFs are
20 preferably engineered so that they contain appropriate cloning sites at their extremities to allow their introduction into a new bacterial expression plasmid, allowing propagation in a standard bacterial host such as *E. coli*, but containing the necessary information for plasmid replication in the target microbe such as *S. aureus* (hereafter referred to as shuttle vector). Shuttle vectors and their use are well known in the art.

25 Such shuttle vectors preferably also contain regulatory sequences that allow inducible expression of the introduced ORF. As the candidate ORF may encode an inhibitor function that will eliminate the host, it is beneficial that it not be expressed prior to testing for activity. Thus, screening for such sequences when expressed in a constitutive fashion is less likely to be successful when the inhibitor is lethal. In the
30 exemplary inducible system presented in Figure 1A, 1B, 2, and 7, regulatory sequences from the *ars* operon of *S. aureus* are used to direct individual ORF expression in *S. aureus* (or other bacteria in which the *ars* system is functional). The *ars* operon encodes a series of proteins which normally mediate the extrusion of arsenite and other trivalent oxyanions from the cells when they are exposed to such
35 toxic substances in their environment. The operon encoding this detoxifying mechanism is normally silent and only induced when arsenite-related compounds are

present. (Tauriainen, S. et al. (1997) *App. Env. Microb.*, Vol. 63, No. 11, p. 4456-4461.)

Therefore, individual phage ORFs can be expressed in *S. aureus* in an inducible fashion by adding to the culture medium non-toxic arsenite concentrations during the growth of individual *S. aureus* clones expressing such individual phage ORFs. Toxicity of the phage inhibitor ORF for the host is monitored by reduction or arrest of growth under induction conditions, as measured by optical density in liquid culture or after plating the induced cultures on solid medium. Subsequently, interference of the phage ORF with the host biochemical pathways ultimately leading to reduced or arrested host metabolism can be measured by pulse-chase experiments using radiolabeled precursors of either DNA replication, RNA transcription, or protein synthesis. Similar constructs can be made and used for other bacteria using well-known techniques.

Those skilled in the art are familiar with a variety of other inducible systems which can also be used for the controlled expression of phage ORFs, including, for example, lactose (see *e.g.*, Stratagene's LacSwitchTMII system; La Jolla, CA) and tetracycline-based systems (see, *e.g.* Clontech's Tet On/Tet OffTM system; Palo Alto, CA). The arsenite-inducible system described is further depicted in Figures 1, 2 and 7.

The selection or construction of shuttle vectors and the selection and use of inducible systems are well known and thus other shuttle vectors appropriate for other bacteria can be readily provided by those skilled in the art, *e.g.*, for use in other bacterial species.

Standard methodologies for expressing proteins from constructs, and isolating and manipulating those proteins, for example in cross-linking and affinity chromatography studies, may be found in various commonly available and known laboratory manuals. See, *e.g.*, Current Protocols in Protein Science, John Wiley & Sons, Secaucus, N.J., and Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.

It has been found that certain phage or other viruses inhibit host cells, at least in part, by producing an antisense RNA which binds to and inhibits translation from a bacterial RNA sequence. Thus, in the case of potentially inhibitor RNA transcripts encoded by the phage genome, a strong indicator of a possible inhibitory function is provided by the identification of phage sequence which is the identical to or fully complementary (or with only a small percentage of mismatch, *e.g.*, <10%, preferably less than 5%, most preferably less than 3%, to a bacterial sequence. This approach is convenient in the case of bacteria that have been essentially completely sequenced, as the comparison can be performed by computer using public database information.

The inhibitory effect of the transcript can be confirmed using expression of the phage sequence in a host bacterium. If needed, such inhibitory can also be tested by transfecting the cells with a vector that will transcribe the phage sequence to form RNA in such manner that the RNA produced will not be translated into a polypeptide.

5 Inhibition under such conditions provides a strong indication that the inhibition is due to the transcript rather than to an encoded polypeptide.

In an alternative, the expression of an ORF in a host bacterium is found to be inhibitory, but the inhibition is found to be due to an RNA product of the genomic coding region. For antisense inhibition, the sequence of the bacterial target nucleic acid sequence can be identified by inspection of the phage sequence, and the full

10 sequence of the relevant coding region for the bacterial product can be found from a database of the bacterial genomic sequence or can be isolated by standard techniques (e.g., a clone in a genomic library can be isolated which contains the full bacterial ORF, and then sequenced).

15 In either case, the identification of a target which is inhibited by an RNA transcript produced by a phage provides both the possible inhibition of bacteria naturally containing the same target nucleic acid sequence, as well as the ability to use the target sequence in screening for other types of compounds which will act directly on the target nucleic acid sequence or on a polypeptide product expressed or

20 regulated, at least in part, by the target of the inhibitory phage RNA.

In some cases it will be found that the target of an inhibitory phage RNA or protein has previously been found to be a target of an inhibitory phage RNA or protein has previously been found to be a target for an antibacterial agent. In such cases, the phage inhibitor can still provide useful information if it is found that the

25 phage-encoded product acts at a different site than the previously identified antibacterial agent or inhibitor, i.e., acts at a phage-specific site. For many targets, action at a different site provides highly beneficial characteristics and/or information. For example, an alternate site of inhibitor action can at least partially overcome a resistance mechanism in a bacterium. As an illustration, in many cases, resistance is

30 due, in large part, to altered binding characteristics of the immediate target to the antibacterial agent. The altered binding is due to a structural change which prevents or destabilizes the binding. However, the structural change is frequently quite local, so that compounds which bind at different local sites will be unaffected or affected to a much lesser degree. Indeed, in some cases the local sites will be on a different

35 molecule and so may be completely unaffected by the local structural change creating resistance to the original agent(s). An example of resistance due to altered binding is

provided by methicillin-resistant *Staphylococcus aureus*, in which the resistance is due to an altered penicillin-binding protein.

In other cases, a new site of action can have improved accessibility as compared to a site acted on by a previously identified agent. This can, for example, assist in allowing effective treatment at lower doses, or in allowing access by a larger range of types of compounds, potentially allowing identification of more potential active agents.

Another advantage is that the structural characteristics of a different site of action will lead to identification and/or development of inhibitors with different structures and different pharmacological parameter. This can allow a greater range of possibilities when selecting an antibacterial agent.

Yet further, different sites often produce different inhibitory characteristics in the target organism. This is commonly the case for multi-domain target proteins. Thus, inhibition targeting an alternate site can produce more efficacious action, e.g., faster killing, slower development of resistance, lower numbers of surviving cells, and different secondary effects (for example, different nutrient utilization).

Staphylococcus aureus phage 77

As indicated above, the present invention is concerned, in part, with the use of bacteriophage 77 coding sequences and the encoded polypeptides or RNA transcripts to identify bacterial targets for potential new antibacterial agents.

As described, phage 77 ORFs 17, 19, 43, 102, 104, and 182 have been found to have bacteria inhibiting function. Identification of ORFs 17, 19, 43, 102, 104, and 182 and products from the phage which inhibit the host bacterium both provides an inhibitor compound and allows identification of the bacterial target affected by the phage-encoded inhibitor. Such a target is thus identified as a potential target for development of other antibacterial agents or inhibitors and the use of those targets to inhibit those bacteria. As indicated above, even if such a target is not initially identified in a particular bacterium, such a target can still be identified if a homologous target is identified in another bacterium. Usually, but not necessarily, such another bacterium would be a genetically closely related bacterium. Indeed, in some cases, an inhibitor encoded by phage 77 ORF 17, 19, 43, 102, 104, or 182 can also inhibit such a homologous bacterial cellular component.

Possible bacterial target sequences are described herein by reference to sequence source sites. In preferred embodiments, the sequence encoding the target corresponds

to a *S. aureus* nucleic acid sequence available from numerous sources including *S. aureus* sequences deposited in GenBank, *S. aureus* sequences found in European Patent Application No. 97100110.7 to Human Genome Sciences, Inc. filed January 7, 1997, *S. aureus* sequences available from TIGR at

- 5 <http://www.tigr.org/tldb/mdb/mdb.html>, and *S. aureus* sequences available from the Oklahoma University *S. aureus* sequencing project at the following URL: http://www.genome.ou.edu/staph_new.html. Such possible targets are particularly applicable to *S. aureus* phages 77, 3A, 96, and 44 AHJD.

- 10 The amino acid sequence of a polypeptide target is readily provided by translating the corresponding coding region. For the sake of brevity, the sequences are not reproduced herein. Also, in preferred embodiments, a target sequence corresponds to a *S. aureus* coding sequence corresponding to a sequence listed in Table 15 herein. The listing in Table 15 describes *S. aureus* sequences currently listed with GenBank. Again, for the sake of brevity, the sequences are described by
- 15 reference to the database accession numbers instead of being written out in full herein. In cases where an entry for a coding region is not complete, the complete sequence can be readily obtained by routine methods, e.g., by isolating a clone in a phage host *S. aureus* genomic library, and sequencing the clone insert to provide the relevant coding region. The boundaries of the coding region can be identified by conventional
- 20 sequence analysis and/or by expression in a bacterium in which the endogenous copy of the coding region has been inactivated and using subcloning to identify the functional start and stop codons for the coding region.

Staphylococcus aureus phage 44 AHJD

- 25 The present invention also can utilize the identification of naturally occurring DNA sequence elements within *Staphylococcus aureus* bacteriophage 44AHJD which encode proteins with antimicrobial activity.

- Such identification can utilize bioinformatics identification of specific proteins (ORFs) utilized by *Staphylococcus aureus* bacteriophage 44AHJD during the viral life
- 30 cycle, resulting in a slowing or arrest of growth of the bacterial host, or in death, of the *Staphylococcus aureus* host including lysis of the infected bacteria. Thus, some of the bacteriophage 44AHJD DNA sequences encoding these proteins (ORFs) are predicted to encode antimicrobial functions. Information derived from these DNA sequences and translated ORFs can, in turn, be utilized to develop inhibitory
- 35 compounds by peptidomimetics that can also function as antimicrobials. In addition, the identification of the host bacterial proteins that are targeted and inhibited by the

antimicrobial bacteriophage ORFs can themselves provide novel targets for drug discovery.

The methodology described above is used to identify and characterize DNA sequences from *Staphylococcus* sp. bacteriophage 44 AHJD that have antimicrobial activity. As described in the Examples, the *Staphylococcus aureus* propagating strain (PS 44A), obtained from the Felix d'Herelle Reference Centre (#HER 1101), was used as a host to propagate its phage 44AHJD, also obtained from the Felix d'Herelle Reference Centre (#HER 101). By sequencing, we found that bacteriophage 44AHJD consists of 16,668 bp (Table 16) predicted to encode 73 ORFs greater than 33 amino acids (Tables 17 & 18). Computational analysis of the predicted protein products of *Staphylococcus aureus* bacteriophage 44AHJD identified homologs in public sequence databases as listed in Table 19 and 20, along with the accompanying list of related proteins.

From this analysis, it is apparent that 3 genes (ORF 3, 7, and 8) are related to structural proteins found in other bacteriophages. These include genes predicted to encode a tail protein (ORF 3), an upper collar/connector protein of the phage virion (ORF 7), and a lower collar protein (ORF 8). Bioinformatics has also identified one gene whose product is likely involved in phage DNA synthesis. One gene (ORF 1) shows significant homology to DNA polymerases of a number of bacteriophages, bacteria and fungi, and the product of this gene is likely responsible for replicating the genetic material of bacteriophage 44AHJD. ORF 2 encodes a protein with homology to the *dinC* gene of *Bacillus subtilis* that encodes a protein involved in teichoic acid biosynthesis. Teichoic acid is a polyphosphate polymer found in some, but not all, Gram positive organisms (and not in Gram negative organisms), where it is attached to the peptidoglycan layer. The phage protein may thus be involved in the synthesis of this material for incorporation into the cell wall, allowing enhanced lysis by the phage lysis enzymes or, as many enzymes can function in "reverse reactions", may be involved in its degradation allowing for penetration of the peptidoglycan and phage genome entry into the cell following adsorption. The similarity between *Staphylococcus aureus* bacteriophage 44AHJD and *E. coli* phage T7 indicates that they may share similar mechanisms of replication and growth. Both phages belong to the Podoviridae Family of bacteriophages and are members of the "T7-like" Genus of this Family (Ackermann and DuBow; Vith ICTV Report).

Two genes, ORF 9 and 12, were identified with the potential to encode antimicrobial protein products. The homology alignments are shown in Tables 19 and 20. The predicted product of ORF 9 is related to a class of genes which encodes lysozyme-like functions, enzymes which cleave linkages in the mucopolysaccharide cell wall structure of a variety of micro-organisms, including that from the *Staphylococcus aureus* bacteriophage Twort. ORF 12 of *Staphylococcus aureus* bacteriophage 44AHJD shows homology to a set of lysis proteins from several bacteriophages. These lysis proteins are also referred to as holins, and represent phage-encoded lysis functions required for transit of the phage murein hydrolases (lysozyme) to the periplasm, where it can digest the cell wall and thus lyse the bacterium.

Thus, in particular embodiments, the present invention provides a nucleic acid sequence isolated from *Staphylococcus aureus* bacteriophage 44AHJD comprising at least a portion of one of the genes described above with antimicrobial activity. For example, ORF 1 encodes a DNA polymerase function. This polymerase may utilize host-derived accessory proteins for its activity when replicating the phage template, sequestering such proteins from use by the bacterial polymerase, resulting in inhibition of DNA replication, cell division, and cell growth. Alternatively, ORF 9 directly encodes a polypeptide with antimicrobial activity. ORF 9 is predicted to encode an amidase, a protein known to act as a cell wall degrading enzyme. ORF 12 likely encodes a holin function required for transit of the phage amidase (gene 9 product) to the periplasm. When this type of gene product from Bacillus phage phi 29 (gene 14), was cloned in *Escherichia coli*, cell death ensued (Steiner et al., 1993). Thus, production of proteins from Bacillus phage phi 29 gene 14 in *E. coli* resulted in cell death, whereas production of protein from Bacillus phage phi 29 gene 14 concomitantly with the phi 29 lysozyme or unrelated murein-degrading enzymes led to lysis, suggesting that membrane-bound protein 14 induces a nonspecific lesion in the cytoplasmic membrane (Steiner et al., 1993).

The present invention also provides the use of the *Staphylococcus* bacteriophage 44 AHJD antimicrobial ORFs or ORF products as pharmacological agents, either wholly or in part and derivatives, as well as the use of corresponding peptidomimetics, developed from amino acid or nucleotide sequence knowledge derived from *Staphylococcus* bacteriophage 44 AHJD killer ORFs.

Enterococcus phage 182

Bacteriophage 182 was obtained from the Felix D'Herelle phage collection (Ste. Foy, Quebec) and infects *Enterococcus* sp. Group D. The genome of
5 *Enterococcus* bacteriophage 182 consists of 17,833 bp (Table 21) and is predicted to encode 80 ORFs greater than 33 amino acids (Tables 22 and 23). Computational analysis of the predicted protein products of *Enterococcus* bacteriophage 182 was performed in order to identify protein products related to those deposited in public databases. Bacteriophage 182 protein products which detected sequences with
10 significant sequence similarity in public databases are listed in Table 24 and 26, along with the accompanying list of related proteins.

From this analysis, it is apparent that 5 genes (ORF 001, 004, 007, 009, and 011) are related to structural proteins of several *Bacillus* phages – *Bacillus* bacteriophage PZA, phi-29, and B103. These include genes predicted to encode a tail
15 protein (ORF 001), a head protein (ORF 004), and upper collar protein (ORF 007), a lower collar protein (ORF 009), and a pre-neck appendage protein (ORF 011). Two gene products are predicted to encode genes which direct phage morphogenesis – these are ORF 005 and 019.

Bioinformatics has also identified three genes whose products are likely
20 involved in phage DNA synthesis. One gene, ORF 002 shows significant homology to DNA polymerases of a number of bacteriophages, and the product of this gene is likely responsible for replicating the genetic material of bacteriophage 182. ORF 006 encodes a protein with homology to the encapsidation proteins of several other bacteriophages, including *Bacillus* phage phi-29 (P11014), PZA (P07541), and B103
25 (X99260) and *Streptococcus* phage CP-1 (Z47794). These gene products catalyze the *in vivo* and *in vitro* genome-encapsidation reaction (Garvey et al., 1985). Proteins involved in genome packaging have been shown to have additional activities that affect biochemical reactions in other phages and their hosts. For example, the coat protein of the RNA bacteriophage MS2 interacts with viral RNA to translationally
30 repress replicase synthesis (Pickett and Peabody, 1993). This protein-RNA interaction also plays a role in genome encapsidation, enveloping a single copy of the viral genome in a protein shell composed of many molecules of coat protein. In addition, the bacteriophage λ terminase enzyme can be lethal to *E. coli* when expressed,

suggesting cleavage of packaging sites in the bacterial chromosome. Also present within bacteriophage 182 is a gene, ORF 010, that encodes a protein that is related to the terminal proteins of *Bacillus* phage Nf (P06812), *Bacillus* phage GA-1 (X96987) and *Bacillus* phage B103 (X99260). DNA terminal proteins are linked to the 5' ends of both strands of the genome and are essential for DNA replication playing a role in initial priming of DNA replication. The similarity between *Enterococcus* bacteriophage 182 and *Bacillus* phages phi-29, PZA, and B103 indicates that they may share similar mechanisms of replication and growth. Protein-primed DNA replication is a well described phenomenon, and in the phi-29-like phages, the ends of the DNA serve as origins and termini of replication (Gutiérrez et al., 1986; Yoshikawa et al., 1985).

There is also a gene (ORF 015) that encodes a protein showing homology to an early protein product of *Bacillus* bacteriophage PZA and the single-strand nucleic acid binding protein of bacteriophage B103.

Two genes, ORF 008 and 014, were identified with the potential to encode anti-microbial protein products. The homology alignments are shown in Tables 24 & 26 and biochemical features of the predicted polypeptides shown in Table 25. The predicted product of ORF 008 is related to a class of genes which encodes lysozyme-like functions, enzymes which cleave linkages in the mucopolysaccharide cell wall structure of a variety of micro-organisms. ORF 014 of *Enterococcus* 182 shows homology to a set of lysis proteins from *Bacillus* bacteriophage phi-29, PZA, and B103. These lysis proteins are also referred to as holins and represent phage encoded lysis functions required for transit of the phage murein hydrolases (lysozyme) to the periplasm, where it can digest the outer cell wall and thus lyse the bacterium.

Thus, the present invention provides a nucleic acid sequence obtained from *Enterococcus* bacteriophage 182 comprising at least a portion of a phage 182 ORF, preferably an inhibitory ORF, and more preferably at least a portion of one of the genes described above with anti-microbial activity. For example, ORF 002 encodes a DNA polymerase function. This polymerase may utilize host-derived accessory proteins for its activity when replicating the phage template, sequestering such proteins from use by the bacterial polymerase, resulting in inhibition of DNA replication, cell division, and cell growth. Alternatively, ORFs 008 or 014 directly encode polypeptides with anti-microbial activity. ORF 008 is predicted to encode an

autolytic lysozyme, a protein known to have anti-microbial activity (Martin *et al.*, 1998). ORF 014 likely encodes a holin function required for transit of the phage murein hydrolases to the periplasm. When the related product from *Bacillus* phage phi 29 (gene 14), was cloned in *Escherichia coli*, cell death ensued (Steiner *et al.*, 1993).

5 Thus, production of proteins from *Bacillus* phage phi 29 gene 14 in *E. coli* resulted in cell death, whereas production of protein from *Bacillus* phage phi 29 gene 14 concomitantly with the phi 29 lysozyme or unrelated murein-degrading enzymes led to lysis, suggesting that membrane-bound protein 14 induces a nonspecific lesion in the cytoplasmic membrane (Steiner *et al.*, 1993).

10 The present invention also provides the use of the *Enterococcus* bacteriophage 182 anti-microbial ORFs as pharmacological agents, either wholly or in part and derivatives, as well as the use of corresponding peptidomimetics, developed from amino acid or nucleotide sequence knowledge derived from *Enterococcus* bacteriophage 182 killer ORFs. This can be done where the structure of the
15 peptidomimetic compound corresponds to the structure of the active portion of a product of an ORF. In this analysis, the peptide backbone is transformed into a carbon based hydrophobic structure that can retain cytostatic or cytotoxic activity for the bacterium. This is done by standard medicinal chemistry methods, measuring growth inhibition of the various molecules in liquid cultures or on solid medium. These
20 mimetics also represent lead compounds for the development of novel antibiotics. In this context, "corresponds" means that the peptidomimetic compound structure has sufficient similarities to the structure of the active portion of a product of one of the *Enterococcus* ORFs listed, that the peptidomimetic will interact with the same molecule as the product of the ORF, and preferably will elicit at least one cellular
25 response in common which relates to the inhibition of the cell by the phage protein.

To validate the identity of an ORF as a killer ORF, it is preferably expressed in the host or other test bacterial organism and the effect of this expression on bacterial growth and replication is assessed. Therefore, all individual ORFs identified herein, e.g., those identified above, can be expressed, preferably overexpressed, in a
30 suitable host bacterium e.g., a host *Enterococcus* and the effect of this expression or overexpression on host metabolism and viability can be measured.

Individual ORFs can be resynthesized from the phage genomic DNA by the polymerase chain reaction (PCR) using oligonucleotide primers flanking the ORF on

either side. Those skilled in the art are familiar with the design and synthesis of appropriate primer sequences. These single ORFs are preferably engineered so that they contain appropriate cloning sites at their extremities to allow their introduction into a new bacterial expression plasmid, allowing propagation in a standard bacterial host such as *E. coli*, but containing the necessary information for plasmid replication in the target microbe, *Enterococcus* sp. (hereafter referred to as a shuttle vector).

This shuttle vector also preferably contains regulatory sequences that allow inducible expression of the introduced ORF. As the candidate ORF may encode a killer function that will eliminate the host, it is highly advantageous that it not be expressed (or at least not expressed at a substantial level) prior to testing for activity; thus screening for such sequences in a constitutive fashion is less likely to be successful (lethality). In an example presented in Fig. 7, regulatory sequences from the *ars* operon are used to direct individual ORF expression in *Enterococcus*. The *ars* operon encodes a series of proteins which normally mediate the extrusion of arsenite and several other trivalent oxyanions from the cells when they are exposed to such toxic substances in their environment. The operon encoding this detoxifying mechanism is normally silent and only induced when arsenite-related compounds are present.

Therefore, individual phage ORFs can be expressed in *Enterococcus* or other suitable host in an inducible fashion by adding to the culture medium non-toxic arsenite concentrations during the growth of individual *Enterococcus* (or other host cells) clones expressing such individual phage ORFs. Toxicity of the phage killer ORF for the host is monitored by reduction or arrest of growth under induction conditions, as measured by optical density in liquid culture or after plating the induced cultures on solid medium. Subsequently, interference of the phage ORF with the host biochemical pathways ultimately leading to reducing or arresting host metabolism can be measured by pulse chase experiments using radiolabeled precursors of either DNA replication, RNA transcription, or protein synthesis.

Of course, other inducible regulatory sequences (e.g., promoters, operators, etc.) may be used (e.g., systems using positive induction of expression or systems using release of repression). A variety of such systems are known to those skilled in the art and can be utilized in the present invention.

Nucleic acid sequences of the present invention can be isolated using a method similar to those described herein or other methods known to those skilled in the art. In addition, such nucleic acid sequences can be chemically synthesized by well-known methods. Having the phage 182 ORFs, e.g., anti-bacterial ORFs of the present invention, portions thereof, or oligonucleotides derived therefrom as described, other anti-microbial sequences from other bacteriophage sources can be identified and isolated using methods described here or other methods, including methods utilizing nucleic acid hybridization and/or computer-based sequence alignment methods.

The invention also provides bacteriophage anti-microbial DNA segments from other phages based on nucleic acids and sequences hybridizing to the presently identified inhibitory ORF under high stringency conditions or sequences which are highly homologous. The bacteriophage anti-microbial DNA segment from bacteriophage 182 can be used to identify a related segment from another unrelated phage based on stringent conditions of hybridization or on being a homolog based on nucleic acid and/or amino acid sequence comparisons. As with the phage 182 inhibitory sequences, such homologous coding sequences and products can be used as antimicrobials, to construct active portions or derivatives, to construct peptidomimetics, and to identify bacterial targets.

Enterococcus sequences are listed in Table 27 by accession number, providing identification of possible targets of *Enterococcus* phage inhibitory ORF products, e.g., from phage 182.

Streptococcus pneumoniae

As indicated in the Summary above, the present invention is concerned with the use of *Streptococcus* sp. bacteriophage Dp-1 coding sequences and the encoded polypeptides or RNA transcripts to identify bacterial targets for potential new antibacterial agents.

Streptococcus pneumoniae is an important cause of community-acquired pneumonia and a major cause of otitis media, sinusitis, and meningitis in children and adults. In Spain and other Mediterranean countries, the majority of *S. pneumoniae* are relatively resistant to penicillin (Klugman, 1990; Fenoli et al., 1991; Jørgensen et al., 1990). These strains also have decreased susceptibility to broad-spectrum cephalosporins, which are frequently used in the empiric treatment of meningitis and

other serious invasive bacterial infections. High-level resistance of pneumococci has been encountered in Hungary where 70% of children who were colonized with *S. pneumoniae* carried penicillin resistant strains that were also resistant to tetracycline, erythromycin, trimethoprim/sulfamethoxazole, and 30% resistant to chloramphenicol (Neu, 1992). The resistance of pneumococci to macrolides such as erythromycin averages 20-25% in France, ~20% in Japan, and <10% in Spain (Neu, 1992).

The antimicrobial susceptibilities and distribution of serotypes of the 42 isolates of *S. pneumoniae* in southern Taiwan from invasive infections have been recently determined (Hseuh et al., 1996). Resistance rates among these isolates were: erythromycin, 61.9%; clindamycin, 47.6%; chloramphenicol, 19%; and tetracycline, 73.8%. Resistance to three or more classes of antibiotics was found in 33.3% of the isolates. Bacteremic pneumonia and primary bacteremia accounted for 64.3% of the infections and mortality was 42.6%. Given the severity of these infections despite adequate antibiotic therapy, there is clearly a need for introduction of new therapeutic options to prevent mortality due to invasive *S. pneumoniae* infections.

Pneumococcal phages belong to four families and they present a great variety in morphology, including lytic and temperate phages (for a review, see Garcia et al., 1997). Examples of lytic phages are Cp-1 and Dp-1, whereas examples of temperate phages are HB-3, EJ-1, and HB-746. The complete nucleotide sequence and functional organization of Cp-1 has been reported (Martin et al., 1996). Cp-1 has a 19,345 bp double-stranded DNA genome, with a terminal protein covalently linked to its 5' ends, that replicates by a protein primed mechanism. The phage contains 29 ORFs, 23 on one strand and 6 on the opposite. When these predicted proteins were compared to sequences compiled in GenBank EMBL databases, to ORFs showed significant similarity to proteins of bacteriophage 29 that infects *B. subtilis* (Martin et al., 1996). The similar proteins corresponded to those involved in DNA replication (terminal protein and DNA polymerase), structural and morphogenic proteins (major head, collar, connector, tail, and encapsidation proteins), and proteins involved in lysis function (holin and lysozyme). In its strategy of lysis, the holin gene product inserts itself into the cell membrane, allowing access of the lysozyme to the peptidoglycan. Expression of the Cp-1 holin protein in *E. coli* results in cell death after 2-hours of induction, but did not lead to lysis (Garcia et al., 1997). Cells harboring a plasmid construction with holin and lysozyme genes together did lyse after induction and the

viability loss was similar to that of the culture expressing holin alone. Cloning of these lytic genes in *S. pneumoniae* showed that both genes had the same effect as in *E. coli*. That is, holin itself did not lyse the culture but the viability loss was noticeable, whereas both holin and lysozyme together were capable of lysing M31, an amidase
5 deleted mutant (Garcia et al., 1997).

Recently, a small portion (~4 kbp) of a second *S. pneumoniae* phage, Dp-1, has been sequenced (Sheehan et al., 1997). This portion contains the genes coding for the lytic system (Sheehan et al., 1997) and shows a modular organization similar to that described for Cp-1. However, in this case, a single chimeric protein appears to be
10 made in which the N-terminal domain is highly similar to that of the murein hydrolase coded by a gene found in the phage BK5-T that infects *Lactococcus lactis*, and the C-terminal domain is homologous to holins. Thus, both functions appear to have been combined in a novel chimeric protein.

Bacteriophage Dp-1 was obtained from Dr. P. Garcia (Departamento de
15 Microbiologia Molecular, Centro de Departamento de Investigaciones Biologicas, Consejo Superior de Investigaciones Cientificas, Velazquez, Madrid, Spain). We found that Dp-1 has a double-stranded DNA genome of 56,506 bp, predicted to encode 85 ORFs greater than 33 amino acids and with upstream Shine-Dalgarno motifs for translation initiation (Tables 28 & 30, and Fig. 6). Computational analysis
20 of the predicted protein products of *Streptococcus* bacteriophage Dp-1 protein products, which detected homologs in public databases, are listed in Table 31, along with the accompanying list of related proteins.

From this analysis, it is apparent that several predicted genes of Dp-1 encode polypeptides that are related to structural proteins. ORFs 001, 002, 004, and 030 are
25 predicted to encode tail proteins, minor structural proteins, and minor capsid proteins (Table 31). We also note the identification of several gene products that are likely involved in DNA synthesis. These include ORF 3 which encodes DNA polymerase, ORF 8 which encodes a SWI/SNF helicase-related protein, ORF 10 encodes a protein showing homology to recA, and ORF 13 encodes a dnaZX-like ORF.

30 In *E. coli*, RapA encodes an RNA polymerase (RNAP)-associated protein with ATPase activity and which is a homolog of the eukaryotic SWI/SNF family, a set of proteins whose members are involved are involved in transcription activation, nucleosome remodeling, and DNA repair. RapA forms a stable complex with RNAP,

as if it were a subunit of RNAP and it is possible that the ORF 8 product behaves similarly or in a dominant-negative fashion to inhibit the activity of RapA. Mutation of the essential *E. coli* dnaZX results in a block in DNA chain elongation during replication (Maki et al., 1988). The dnaZX gene has only one open reading frame for a 71-kDa polypeptide from which the two distinct DNA polymerase III holoenzyme subunits, tau (71 kDa) and gamma (47 kDa), are produced. The tau subunit is the precursor of the gamma subunit, and the gamma subunit is produced by a -1 frameshift causing early termination of translation (Tsuchihashi et al., 1990). These proteins show single-strand DNA binding properties that is ATPase (and dATPase) dependent and are thought to increasing the processivity of the core DNA polymerase enzyme (Lee et al., 1987).

There are several Dp-1 ORFs which encode proteins predicted to play a role in cellular metabolic pathways. These include polypeptides involved in coenzyme PQQ synthesis (ORFs 20, 29, 38). Pyrrolo-quinoline quinone (PQQ) is the non-covalently bound prosthetic group of many quinoproteins catalysing reactions in the periplasm of Gram-negative bacteria. Most of these involve the oxidation of alcohols or aldose sugars. Interestingly, ORFs 20, 29, and 30 also show homology to the exoenzyme S regulon (Frank, 1997). Proteins encoded by the *P. aeruginosa* exoenzyme S regulon may be involved in a contact-mediated translocation mechanism to transfer anti-host factors directly into eukaryotic cells disrupting eukaryotic signal transduction through ADP-ribosylation (Frank, 1997).

There is also a protein with similarity to GTP cyclohydrolase I (ORF 21) and ORF 41 which shows homology to dUTPase (Table 31). GTP cyclohydrolase I is an enzyme that catalyzes the first reaction in the pathway for the biosynthesis of the pteridine, a cofactor of the monooxygenases of the aromatic amino acids. Disruption of the homologous gene in *Saccharomyces cerevisiae* leads to a recessive conditional lethality due to folinic acid auxotrophy, that can be complemented with the mammalian or bacterial GTP cyclohydrolase I enzymes (Nardese et al., 1996; Mancini et al., 1999).

ORF 16 shows high homology to autolysin. This region of the phage sequence was previously reported (Sheehan et al., 1997) and encompasses ~ 4 kbp of our sequence. The sequence published by (Sheehan et al., 1997) is shown in Table 32.

Thus, the present invention provides a nucleic acid sequence obtained from *Streptococcus* bacteriophage Dp-1 comprising at least a portion of a phage Dp-1 ORF; preferably an inhibitory ORF, and more preferably at least a portion of one of the genes described above with anti-microbial activity. For example, ORF 013 encodes a

protein with homology to the gamma subunit of DNA polymerase (dnaX gene). This protein may act in a dominant-negative fashion to sequester the host DNA polymerase for its own replication, thus inhibiting host DNA replication. The dnaX gene product is essential for *E. coli* replication (Kodaira et al., 1983).

5 In certain preferred embodiments of the present invention, the bacterial target of a bacteriophage inhibitor ORF product, e.g., an inhibitory protein or polypeptide, is encoded by a *Streptococcus* nucleic acid coding sequence from a host bacterium for bacteriophage Dp-1. As above, possible target sequences are described herein by reference to sequence source sites. The sequence encoding the target preferably
10 corresponds to a *Streptococcus* nucleic acid sequence available from The Institute for Genomic Research (TIGR), or available from GenBank or other public database. The TIGR *Streptococcus* sequences are publicly available at The Institute for Genomics Research at URL: <http://www.tigr.org>

The amino acid sequence of a polypeptide target is readily provided by
15 translating the corresponding coding region. For the sake of brevity, the sequences are not reproduced herein. Also, in preferred embodiments, a target sequence corresponds to a *Streptococcus pneumoniae* coding sequences corresponding to a sequence listed in Table 33 herein. Sequences for other Streptococcal species are also available from TIGR and/or from GenBank. The listing in Table 33 describes
20 *Streptococcus* sequences currently deposited in GenBank. Again, for the sake of brevity, the sequences are described by reference to the GenBank entries instead of being written out in full herein. In cases where the TIGR or GenBank entry for a coding region is not complete, the complete sequence can be readily obtained by routine methods, e.g., by isolating a clone in a phage Dp-1 host *Streptococcus* sp.
25 genomic library, and sequencing the clone insert to provide the relevant coding region. The boundaries of the coding region can be identified by conventional sequence analysis and/or by expression in a bacterium in which the endogenous copy of the coding region has been inactivated and using subcloning to identify the functional start and stop codons for the coding region.

30 In the various aspects of this invention involving Dp-1 sequences, preferably the sequence is preferably not contained in the sequence described in Sheehan et al., 1997 (Table 32).

Validating Identified Inhibitory Phage ORFs

35 A fifth step involves validating the identified phage inhibitor ORF by independent methods, and delineating further possible smaller segments of the ORFs

that have inhibitory activity. Several methods exist to validate the role of the identified ORF as an inhibitor ORF.

One example utilizes the creation of a mutant variant of the phage ORF in which the candidate ORF carries a partial or complete loss-of-function mutation that is measurable as compared with the non-mutant ORF. Comparison of the effects of expression of the loss of function mutant with the normal ORF provides confirmation of the identification of an inhibitor ORF where the loss-of-function mutant provides a measurably lower level of inhibition, preferably no inhibition. The loss of function may be conditional, *e.g.*, temperature sensitive.

Once validation of the inhibitor ORF is achieved, a bi-directional deletion analysis can be carried out using the same experimental system to identify the minimal polypeptide segment that has inhibitor activity. This may be carried out by a variety of means, *e.g.*, by exonuclease or PCR methodologies, and is used to determine if a relatively small segment of the ORF (*i.e.*, the product of the ORF) still possesses inhibitory activity when isolated away from its native sequence. If so, a portion of the ORF encoding this "active portion" can be used as a template for the synthesis of novel anti-microbial agents and further allowing derivation of the peptide sequence, *e.g.*, using modified peptides and/or peptidomimetics.

In creation of certain peptidomimetics, the peptide backbone is transformed into a carbon-based hydrophobic structure that can retain inhibitor activity against the bacterium. This is done by standard medicinal chemistry methods, typically monitored by measuring growth inhibition of the various molecules in liquid cultures or on solid medium. These mimetics can also represent lead compounds for the development of novel antibiotics.

Recently, a major effort has been undertaken by the pharmaceutical industry and their biotechnology partners for the sequencing of bacterial pathogen genomes. The rationale is that the systematic sequencing of the genome will identify all of the bacterial proteins and therefore this proteome will be the target for designing novel inhibitor antibiotics. Although systematic, this approach has several major problems. The first is that analysis of primary amino acid sequences of bacterial proteins does not immediately reveal which protein will be essential for viability of the bacterium, and target validation is thus a major issue. The second problem is one of redundancy, as several biochemical pathways are either structurally duplicated in bacteria (different isoforms of the same enzyme), or functionally duplicated by the presence of salvage pathways in the event of a metabolic block in one pathway (different nutritional conditions). The third is that even a valid target may not be structurally or

functionally amenable to inhibition by small molecules because of inaccessibility (sequestration of target).

Therefore, there is considerable interest within the pharmaceutical and biotechnology industry in identifying key targets for drug discovery amongst the mass
5 of novel targets generated by large-scale genomic sequencing projects.

On the other hand, and underscoring the instant invention, the phages herein described have, over millions of years, evolved specific mechanisms to target such key biochemical pathways and proteins. In the few cases where inhibition by phages has been elucidated (*e.g.*, see ref. 3), such bacterial targets are invariably rate-limiting
10 in their respective biochemical pathways, are not redundant, and/or are readily accessible for inhibition by the phage (or by another inhibitory compound). Therefore, the sixth step of this invention involves identifying the host biochemical pathways and proteins that are targeted by the phage inhibitory mechanisms.

15 Identifying, Validating, and Characterizing Bacterial Host Target Proteins and Affected Pathways

A rationale for this step is that the inhibitor ORF product from the phage physically interacts with and/or modifies certain microbial host components to block their function. Exemplary approaches which can be used to identify the host bacterial
20 pathways and proteins that interact with, and preferably also are inhibited by, phage ORF product(s) are described below.

One approach is a genetic screen to determine physiological protein:protein interaction, for example, using a yeast two hybrid system. In this assay, the phage ORF is fused to the carboxyl terminus of the yeast Gal4 activation domain II (amino
25 acids 768-881) to create a bait vector. A cDNA library of cloned *S. aureus* sequences which have been engineered into a plasmid where the *S. aureus* sequences are fused to the DNA binding domain of Gal4 is also generated. These plasmids are introduced alone, or in combination, into yeast strain Y190 - previously engineered with chromosomally integrated copies of the *E. coli lacZ* and the selectable HIS3 genes,
30 both under Gal4 regulation (Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., and Elledge, S.J. (1993). *Genes & Dev.* 7, 555-569). If the two proteins expressed in yeast interact, the resulting complex will activate transcription from promoters containing Gal4 binding sites. A *lacZ* and His3 gene, each driven by a promoter containing Gal4 binding sites, have been integrated into the
35 genome of the host yeast system used for measuring protein-protein interactions. Such a system provides a physiological environment in which to detect potential protein interactions. This system has been extensively used to identify novel protein-protein

interaction partners and to map the sites required for interaction (for example, to identify interacting partners of translation factors (Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998). *Mol & Cell Biology* 18, 2697-2711), transcription factors (Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura, Y., and Miki, Y. (1998). *Genes, Chromosomes & Cancer* 21, 217-222), and proteins involved in signal transduction (Endo, T.A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. *Nature* 387, 921-924). This approach has also been used in many published reports to identify interaction between mammalian viral and mammalian cell proteins.

For example, the non-structural protein NS1 of parvovirus is essential for viral DNA amplification and gene expression and is also the major cytopathic effector of these viruses. A yeast two-hybrid screen with NS1 identified a novel cellular protein of unknown function that interacts with NS-1, called SGT, for small glutamine-rich tetratricopeptide repeat (TPR)-containing protein (Cziepluch C. Kordes E. Poirey R. Grewenig A. Rommelaere, J. and Jauniaux JC. (1998) *J Virol.* 72, 4149-4156). In another screen, the adenovirus E3 protein was recently shown to interact with a novel tumor necrosis factor alpha-inducible protein and to modulate some of the activities of E3 (Li Y. Kang J. and Horwitz M.S. (1998). *Mol & Cell Biol.* 18, 1601-1610). In yet another recent screen, the herpes simplex virus 1 alpha regulatory protein ICP0 was found to interact with (and stabilize) the cell cycle regulator cyclin D3 (Kawaguchi Y. Van Sant C. and Roizman B. (1997). *J Virol.* 71, 7328-7336).

Another two-hybrid system for identifying protein:protein interactions is commercially available from STRATEGENE™ as the CYTO-TRAP™ system (Chang et al., *Strategies Newsletter* 11(3), 65-68 (1998)(from Stratagene)). The system is a yeast-based method for detecting protein:protein interactions *in vivo*, using activation of the Ras signal transduction cascade by localizing a signal pathway component, human Sos (hSos), to its activation site in the yeast plasma membrane. The system uses a temperature-sensitive *Saccharomyces cerevisiae* mutant, strain cdc25H, which contains a point mutation at amino acid residue 1328 of the cdc25 gene. This gene encodes a guanyl nucleotide exchange factor which binds and activates Ras, leading to cell growth. The mutation in the cdc25 gene prevents host growth at 37°C, but at a permissive temperature of 25°C, growth is normal. The system utilizes the ability of (hSos) to complement the cdc25 defect and activate the yeast Ras signaling pathway. Once (hSos) is expressed and localized to the plasma membrane, the cdc25H yeast strain grows at 37°C. Localizing hSos to the plasma

membrane occurs through a protein:protein interaction. A protein of interest, or bait, is expressed as a fusion protein with hSos. The library, or target proteins are expressed with the myristylation membrane-localization signal. The yeast cells are then incubated under restrictive conditions (37°C). If the bait and the target protein interact, the hSos protein is recruited to the membrane, activating the Ras signaling pathway and allowing the cdc25H yeast strain to grow at the restrictive temperature.

5 The protein targets of phage inhibitory ORFs can also be identified using bacterial genetic screens. One approach involves the overexpression of a phage inhibitory protein in mutagenized bacterial host species, followed by plating the cells and searching for colonies that can survive the antimicrobial activity of the inhibitory ORF. These colonies are then grown, their DNA extracted, and cloned into an expression vector that contains a replicon of a different incompatibility group from the plasmid expressing the original ORF. This library is then introduced into a wild-type host bacterium in conjunction with an expression vector driving synthesis of the phage ORF, followed by selection for surviving bacteria. Thus, bacterial DNA fragments from the survivors presumably contain a DNA fragment from the original mutagenized host bacterial genome that can protect the cell from the antimicrobial activity of the inhibitory phage ORF. This fragment can be sequenced and compared with that of the bacterial host to determine in which gene the mutation lies. This approach enables one to determine the targets and pathways that are affected by the killing function.

A second approach is based on identifying protein:protein interactions between the phage ORF product and bacterial *S. aureus*, e.g., proteins using a biochemical approach based, for example, on affinity chromatography. This approach has been used, for example, to identify interactions between lambda phage proteins and proteins from their *E. coli* host (Sopta, M., Carthew, R.W., and Greenblatt, J. (1985) *J. Biol. Chem.* 260, 10353-10369). The phage ORF is fused to a peptide tag (e.g. glutathione-S-transferase ("GST"), 6xHIS, ("HIS") and/or calmodulin binding protein ("CPB")) within a commercially available plasmid vector that directs high level expression on induction of a suitably responsive promoter driving the fusion's expression. The translated fusion protein is expressed in *E. coli*, purified, and immobilized on a solid phase matrix via, for example the tag. Total cell extracts from the host bacterium, e.g., *S. aureus*, are then passed through the affinity matrix containing the immobilized phage ORF fusion protein; host proteins retained on the column are then eluted under different conditions of ionic strength, pH, detergents etc., and characterized by gel electrophoresis and other techniques. Appropriate controls are run to guard against nonspecific binding to the resin. Target proteins thus

recovered should be enriched for the phage protein/peptide of interest and are subsequently electrophoretically or otherwise separated, purified, sequenced, or biochemically analyzed. Usually sequencing entails individual digestion of the proteins to completion with a protease (e.g., trypsin), followed by molecular mass and amino acid composition and sequence determination using, for example, mass spectrometry, e.g., by MALDI-TOF technology (Qin, J., Fenyó, D., Zhao, Y., Hall, W.W., Chao, D.M., Wilson, C.J., Young, R.A. and Chait, B.T. (1997). *Anal. Chem.* 69, 3995-4001).

The sequence of the individual peptides from a single protein are then analyzed by the bioinformatics approach described above to identify the *S. aureus* protein interacting with the phage ORF. This analysis is performed by a computer search of the *S. aureus* genome for an identified sequence. Alternatively, all tryptic peptide fragments of the *S. aureus* genome can be predicted by computer software, and the molecular mass of such fragments compared to the molecular mass of the peptides obtained from each interacting protein eluted from the affinity matrix. The responsible gene sequence can be obtained, for example by using synthetic degenerate nucleic acid sequences to pull out the corresponding homologous bacterial sequence. Alternatively, antibodies can be generated against the peptide and used to isolate nascent peptide/mRNA transcript complexes, from which the mRNA can be reverse transcribed, cloned, and further characterized using the procedures discussed herein.

A variety of other binding assay methods are known in the art and can be used to identify interactions between phage proteins and bacterial proteins or other bacterial cell components. Such methods that allow or provide identification of the bacterial component can be used in this invention for identifying putative targets.

Validation of the interaction between the phage ORF product and the bacterial proteins or other components can be obtained by a second independent assay (e.g., co-immunoprecipitation or protein-protein crosslinking experiments (Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998). *Mol & Cell Biology* 18, 2697-2711; Brown, S. and Blumenthal, T. (1976). *Proc. Natl. Acad. Sci. USA* 73, 1131-1135)).

Finally, the essential nature of the identified bacterial proteins is preferably determined genetically by creating a constitutive or inducible partial or complete loss-of-function mutation in the gene encoding the identified interacting bacterial protein. This mutant is then tested for bacterial survival and replication.

The protein target of the phage inhibitor function can also be identified using a genetic approach. Two exemplary approaches will be delineated here. The first approach involves the overexpression of a predetermined phage inhibitor protein in mutagenized host bacteria, e.g., *S. aureus*, followed by plating the cells and searching

for colonies that can survive the inhibitor. These colonies will then be grown, their DNA extracted and cloned into an expression vector that contains a replicon of a different incompatibility group, and preferably having a different selectable marker than the plasmid expressing the phage inhibitor. Thus, host DNA fragments from the mutant that can protect the cell from phage ORF inhibition can be sequenced and compared with that of the bacterial host to determine in which gene the mutation lies. This approach allows rapid determination of the targets and pathways that are affected by the inhibitor.

Alternatively, the bacterial targets can be determined in the absence of selecting for mutations using an approach known as "multicopy suppression". In this approach, the DNA from the wild type host is cloned into an expression vector that can coexist, as previously described, with one containing a predetermined phage inhibitor. Those plasmids that contain host DNA fragments and genes that protect the host from the phage inhibitor can then be isolated and sequenced to identify putative targets and pathways in the host bacteria.

Regardless of the specific mode of identification, screening assays may additionally utilize gene fusions to specific "reporter genes" to identify a bacterial gene(s) whose expression is affected when the host target pathway is affected by the phage inhibitor. Such gene fusions can be used to search a number of small molecule compounds for inhibitors that may affect this pathway and thus cause cell inhibition. This approach will allow the screening of a large number of molecules on petri dishes or 96-well format by monitoring for a simple color change in the bacterial colonies. In this manner, we can validate host targets and classes of compounds for further study and clinical development. These inhibitors also represent lead compounds for the development of other antibiotics.

Bioinformatics and comparative genomics are preferably then applied to the identified bacterial gene products to predict biochemical function. The biochemical activity of the protein can be verified *in vitro* in cell free assays or *in vivo* in intact cells. *In vitro* biochemical assays utilizing cell-free extracts or purified protein are established as a basis for the screening and development of inhibitors.

These inhibitors, preferably small molecule inhibitors, may comprise peptides, antibodies, products from natural sources such as fungal or plant extracts or small molecule organic compounds. In general, small molecule organic compounds are preferred. These compounds may, for example, be identified within large compound libraries, including combinatorial libraries. For example, a plurality of compounds, preferably a large number of compounds can be screened to determine whether any of the compounds binds or otherwise disrupts or inhibits the identified bacterial target.

Compounds identified as having any of these activities can then be evaluated further in cell culture and/or animal model systems to determine the pharmacological properties of the compound, including the specific anti-microbial ability of the compound.

- 5 For mixtures of natural products, including crude preparations, once a preparation or fraction of a preparation is shown to have an anti-microbial activity, the active substance can be isolated and identified using techniques well known in the art, if the compound is not already available in a purified form.

- 10 Identified compounds possessing anti-microbial activity and similar compounds having structural similarity can be further evaluated and, if necessary, derivatized according to synthesis and/or modification methods available in the art selected as appropriate for the particular starting molecule.

Derivatization of identified anti-microbials

- 15 In cases where the identified anti-microbials above might represent peptidal compounds, the *in vivo* effectiveness of such compounds may be advantageously enhanced by chemical modification using the natural polypeptide as a starting point and incorporating changes that provide advantages for use, for example, increased stability to proteolytic degradation, reduced antigenicity, improved tissue penetration, and/or improved delivery characteristics.

- 20 In addition to active modifications and derivative creations, it can also be useful to provide inactive modifications or derivatives for use as negative controls or introduction of immunologic tolerance. For example, a biologically inactive derivative which has essentially the same epitopes as the corresponding natural antimicrobial can be used to induce immunological tolerance in a patient being treated. The induction of tolerance can then allow uninterrupted treatment with the active anti-microbial to continue for a significantly longer period of time.

- 25 Modified anti-microbial polypeptides and derivatives can be produced using a number of different types of modifications to the amino acid chain. Many such methods are known to those skilled in the art. The changes can include, for example, reduction of the size of the molecule, and/or the modification of the amino acid sequence of the molecule. In addition, a variety of different chemical modifications of the naturally occurring polypeptide can be used, either with or without modifications to the amino acid sequence or size of the molecule. Such chemical modifications can, for example, include the incorporation of modified or non-natural amino acids or non-amino acid moieties during synthesis of the peptide chain, or the post-synthesis modification of incorporated chain moieties.

The oligopeptides of this invention can be synthesized chemically or through an appropriate gene expression system. Synthetic peptides can include both naturally occurring amino acids and laboratory synthesized, modified amino acids.

Also provided herein are functional derivatives of anti-microbial proteins or polypeptides. By "functional derivative" is meant a "chemical derivative,"
5 "fragment," "variant," "chimera," or "hybrid" of the polypeptide or protein, which terms are defined below. A functional derivative retains at least a portion of the function of the protein, for example reactivity with a specific antibody, enzymatic activity or binding activity.

10 A "chemical derivative" of the complex contains additional chemical moieties not normally a part of the protein or peptide. Such moieties may improve the molecule's solubility, absorption, biological half-life, and the like. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, and the like. Moieties capable of mediating
15 such effects are disclosed in Alfonso and Gennaro (1995). Procedures for coupling such moieties to a molecule are well known in the art. Covalent modifications of the protein or peptides are included within the scope of this invention. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting
20 with selected side chains or terminal residues, as described below.

Cysteinyl residues most commonly are reacted with alpha-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate, N-
25 alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloro-mercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-
30 bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysiny and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysiny residues. Other suitable reagents for derivatizing
35 primary amine-containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride;

trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pK_a of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine alpha-amino group.

Tyrosyl residues are well-known targets of modification for introduction of spectral labels by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction carbodiimide ($R'-N-C-N-R'$) such as 1-cyclohexyl-3-(2-morpholinyl(4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Glutaminyl and asparaginyl residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Derivatization with bifunctional agents is useful, for example, for cross-linking component peptides to each other or the complex to a water-insoluble support matrix or to other macromolecular carriers. Commonly used cross-linking agents include, for example, 1,1-bis (diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl) dithiolpropioimide yield photoactivatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (Creighton, T.E.,

Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine, and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the stability, solubility, absorption, biological half life, and the like. The moieties may alternatively eliminate or attenuate any undesirable side effect of the protein complex. Moieties capable of mediating such effects are disclosed, for example, in Alfonso and Gennaro (1995).

The term "fragment" is used to indicate a polypeptide derived from the amino acid sequence of the protein or polypeptide having a length less than the full-length polypeptide from which it has been derived. Such a fragment may, for example, be produced by proteolytic cleavage of the full-length protein. Preferably, the fragment is obtained recombinantly by appropriately modifying the DNA sequence encoding the proteins to delete one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence.

Another functional derivative intended to be within the scope of the present invention is a "variant" polypeptide that either lacks one or more amino acids or contains additional or substituted amino acids relative to the native polypeptide. The variant may be derived from a naturally occurring polypeptide by appropriately modifying the protein DNA coding sequence to add, remove, and/or to modify codons for one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence.

A functional derivative of a protein or polypeptide with deleted, inserted and/or substituted amino acid residues may be prepared using standard techniques well-known to those of ordinary skill in the art. For example, the modified components of the functional derivatives may be produced using site-directed mutagenesis techniques (as exemplified by Adelman et al., 1983, *DNA* 2:183; Sambrook et al., 1989) wherein nucleotides in the DNA coding sequence are modified such that a modified coding sequence is produced, and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host cell, using techniques such as those described above. Alternatively, components of functional derivatives of complexes with amino acid deletions, insertions and/or substitutions may be conveniently prepared by direct chemical synthesis, using methods well-known in the art.

Insofar as other anti-microbial inhibitor compounds identified by the invention described herein may not be peptidal in nature, other chemical techniques exist to allow their suitable modification, as well, and according the desirable principles discussed above.

Administration and Pharmaceutical Compositions

For the therapeutic and prophylactic treatment of infection, the preferred method of preparation or administration of anti-microbial compounds will generally vary depending on the precise identity and nature of the anti-microbial being delivered. Thus, those skilled in the art will understand that administration methods known in the art will also be appropriate for the compounds of this invention.

The particularly desired anti-microbial can be administered to a patient either by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s). In treating an infection, a therapeutically effective amount of an agent or agents is administered. A therapeutically effective dose refers to that amount of the compound that results in amelioration of one or more symptoms of bacterial infection and/or a prolongation of patient survival or patient comfort.

Toxicity, therapeutic and prophylactic efficacy of anti-microbials can be determined by standard pharmaceutical procedures in cell cultures and/or experimental organisms such as animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50} . Compounds that exhibit large therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized.

For any compound identified and used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. Such information can be used to more accurately determine useful doses in organisms such as plants and animals, preferably mammals, and most preferably humans. Levels in plasma may be measured, for example, by HPLC or other means appropriate for detection of the particular compound.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition (see *e.g.* Fingl et. al., in *The Pharmacological Basis of Therapeutics*, 1975, Ch. 1 p.1).

It should be noted that the attending physician would know how and when to terminate, interrupt, or adjust administration due to toxicity, organ dysfunction, or other systemic malady. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate (precluding

toxicity). The magnitude of an administered dose in the management of the disorder of interest will vary with the severity of the condition to be treated and the route of administration. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose
5 frequency, will also vary according to the age, body weight, and response of the individual patient. A program comparable to that discussed above also may be used in veterinary or phyto medicine.

Depending on the specific infection target being treated and the method selected, such agents may be formulated and administered systemically or locally, i.e.,
10 topically. Techniques for formulation and administration may be found in Alfonso and Gennaro (1995). Suitable routes may include, for example, oral, rectal, transdermal, vaginal, transmucosal, intestinal, parenteral, intramuscular, subcutaneous, or intramedullary injections, as well as intrathecal, intravenous, or intraperitoneal injections.

15 For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

20 Use of pharmaceutically acceptable carriers to formulate identified anti-microbials of the present invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular those formulated as solutions, may be administered parenterally, such as by intravenous
25 injection. Appropriate compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

30 Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. Liposomes are spherical lipid bilayers with aqueous interiors. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the
35 aqueous interior. The liposomal contents are both protected from the external microenvironment and, because liposomes fuse with cell membranes, are efficiently

delivered into the cell cytoplasm. Additionally, due to their hydrophobicity, small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tablets, dragees, capsules, or solutions, including those formulated for delayed release or only to be released when the pharmaceutical reaches the small or large intestine.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating, entrapping or lyophilizing processes.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active anti-microbial compounds in water-soluble form. Alternatively, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

- 5 Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

- Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added.

- 15 The above methodologies may be employed either actively or prophylactically against an infection of interest.

Computer-related Aspects and Embodiments

- In addition to the provision of compounds as chemical entities, nucleotide sequences, or fragments thereof at least 95%, preferably at least 97%, more preferably at least 99%, and most preferably at least 99.9% identical to phage inhibitor sequences can also be provided in a variety of additional media to facilitate various uses.

- Thus, as used in this section, "provided" refers to an article of manufacture, rather than an actual nucleic acid molecule, which contains a nucleotide sequence of the present invention; *e.g.*, a nucleotide sequence of an exemplary bacteriophage or a sequence encoding a bacterial target or a fragment thereof, preferably a nucleotide sequence at least 95%, more preferably at least 99% and most preferably at least 99.9% identical to such a bacteriophage or bacterial sequence, for example, to a polynucleotide of an unsequenced phage listed in Table 1, preferably of bacteriophage 77 (*S. aureus* host) or bacteriophage 3A (*S. aureus* host) or bacteriophage 96 (*S. aureus* host). Such an article provides a large portion of the particular bacteriophage genome or bacterial gene and parts thereof (*e.g.*, a bacteriophage open reading frame (ORF)) in a form which allows a skilled artisan to examine and/or analyze the sequence using means not directly applicable to examining the actual genome or gene, or subset thereof as it exists in nature or in purified form as a chemical entity.

In one application of this aspect, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer

readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create an article of manufacture which includes one or more computer readable media having recorded thereon a nucleotide sequence or sequences of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can, for example, be presented in a word processing text file, formatted in commercially available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form a nucleotide sequence of an unsequenced bacteriophage, such as an exemplary bacteriophage listed in Table 1 or of a sequence encoding a bacterial target or a fragment thereof, preferably a nucleotide sequence at least 95%, more preferably at least 99% and most preferably at least 99.9% identical to such a bacteriophage or bacterial sequence, for example, to a polynucleotide of bacteriophage 77 (*S. aureus* host) or bacteriophage 3A (*S. aureus* host) bacteriophage

96 (*S. aureus* host), bacteriophage 44AHJD (*S. aureus* host), bacteriophage Dp-1 (*Streptococcus pneumoniae* host), or bacteriophage 182 (*Enterococcus* host) the present invention enables the skilled artisan to routinely access the provided sequence information for a wide variety of purposes.

5 Those skilled in the art understand that software can implement a variety of different search or analysis software which implement sequence search and analysis algorithms, *e.g.*, the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990) and BLAZE (Brutlag et al., Comp. Chem 17:203-207 (1993)) search algorithms. For example, such search algorithms can be implemented on a Sybase system and used to
10 identify open reading frames (ORFs) within the bacteriophage genome which contain homology to ORFs or proteins from other viruses, *e.g.*, other bacteriophage, and other organisms, *e.g.*, the host bacterium. Among the ORFs discussed herein are protein encoding fragments of the bacteriophage genomes which encode bacteria-inhibiting proteins or fragments.

15 The present invention further provides systems, particularly computer-based systems, which contain the sequence information described. Such systems are designed to identify, among other things, useful fragments of the bacteriophage genomes.

 As used herein, "a computer-based system" refers to the hardware, software,
20 and data storage media used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input device, output device, and data storage medium or media. A skilled artisan will readily recognize that any of the currently available general purpose computer-based system are suitable
25 for use in the present invention, as well as a variety of different specialized or dedicated computer-based systems.

 As stated above, the computer-based systems of the present invention comprise data storage media having stored therein a nucleotide sequence of the present invention and the necessary hardware and software for supporting and
30 implementing a search and/or analysis program.

 As used herein, "data storage media" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

35 As used herein, "search program" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means.

Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches and/or sequence analyses can be adapted for use in the present computer-based systems.

As used herein in connection with sequence searches and analyses, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Also, the target sequence length is preferably selected to include sequence corresponding to a biologically relevant portion of an encoded product, for example a region which is expected to be conserved across a range of source organisms. Preferably the sequence length of a target polypeptide sequence is from 5-100 amino acids, more preferably 7-50 or 7-100 amino acids, and still more preferably 10-80 or 10-100 amino acids. Preferably the sequence length of a target polynucleotide sequence is from 15-300 nucleotide residues, more preferably from 21-240 or 21-300, and still more preferably 30-150 or 30-300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length. Likewise, it may be desirable to search and/or analyze longer sequences.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output devices can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output device ranks fragments of the bacteriophage or bacterial sequences possessing varying degrees of homology to the

target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing methods and/or devices and/or formats can be used to
5 compare a target sequence or target motif with the sequence stored in data storage media to identify sequence fragments of the bacteriophage or bacterium in question. One skilled in the art can readily recognize that any one of the publicly available homology search programs can be used as the search program for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be
10 known to those of skill, or later developed, also may be employed in this regard.

Figure 6 provides a block diagram of a computer system illustrative of
embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety
15 of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, etc. A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, etc.) containing control logic and/or data recorded therein may be inserted into
20 the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well-known
25 manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing and processing the sequence (such as search tools, comparing tools, etc.) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

30 The data storage medium in which the sequence is embodied and the central processor need not be part of a single stand-alone computer, but may be separated so long as data transfer can occur. For example, the processor or processors being utilized for a search or analysis can be part of one general purpose computer, and the data storage medium can be part of a second general purpose computer connected to a
35 network, or the data storage medium can be part of a network server. As another example the data storage medium can be part of a computer system or network accessible over telephone lines or other remote connection method.

EXAMPLES

Example 1. Growth of *Staph A* bacteriophage 77 and purification of genomic DNA.

5 The *Staphylococcus aureus* propagating strain (PS 77; ATCC #27699) was used as a host to propagate its respective phage 77 (ATCC # 27699-B1). Two rounds of plaque purification of phage 77 were performed on soft agar essentially as described in Sambrook et al (1989). Briefly, the PS 77 strain was grown overnight at 37°C in Nutrient broth [NB: 0.3% Bacto beef extract, 0.5% Bacto peptone (Difco
10 Laboratories) and 0.5% NaCl (w/v)]. The culture was then diluted 20x in NB and incubated at 37°C until the $OD_{540} = .2$ (early log phase) with constant agitation. In order to obtain single plaques, phage 77 was subjected to 10-fold serial dilutions using phage buffer (1 mM $MgSO_4$, 5 mM $MgCl_2$, 80 mM NaCl and 0.1% Gelatin (w/v)) and 10 μ l of each dilution was used to infect 0.5 ml of the cell suspension in the presence
15 of 400 μ g/ml $CaCl_2$. After incubation of 15 min at room temperature (RT), 2 ml of melted soft agar kept at 45°C (NB supplemented with 0.6% agar) was added to the mixture and poured onto the surface of 100 mm nutrient agar plates (0.3% Bacto Beef extract, 0.5% Bacto peptone, 0.5% NaCl and 1.5% Bacto agar (w/v)). After overnight incubation at 30°C, a single plaque was isolated, resuspended in 1 ml of phage buffer
20 by end over end rotation for 2 hrs at 20°C, and the phage suspension was diluted and used for a second infection as described above. After overnight incubation at 30°C, a single plaque was isolated and used as a stock.

 The propagation procedure for bacteriophage 77 was modified from the agar layer method of Swanstörn and Adams (1951). Briefly, the PS 77 strain was grown to
25 stationary phase overnight at 37°C in Nutrient broth. The culture was then diluted twenty-fold in NB and incubated at 37°C until the $OD_{540} = .2$. The suspension (15×10^7 Bacteria) was then mixed with 15×10^5 plaque forming units (pfu) to give a ratio of 100-bacteria/phage particle in the presence of 400 μ g/ml of $CaCl_2$. After incubation for 15 min at 20°C, 7.5 ml of melted soft agar (NB plus 0.6% agar) were added to the
30 mixture and poured onto the surface of 150 mm nutrient agar plates and incubated 16 hrs at 30°C. To collect the phage plate lysate, 20 ml of NB were added to each plate and the soft agar layer was collected by scrapping off with a clean microscope slide followed by shaking of the agar suspension for 5 min to break up the agar. The mixture was then centrifuged for 10 min at 4,000 RPM (2,830xg) in a JA-10 rotor
35 (Beckman) and the supernatant fluid (lysate) was collected and subjected to a treatment with 10 μ g /ml of DNase I and RNase A for 30 min at 37°C. To precipitate the phage particles, the phage suspension was adjusted to 10% (w/v) PEG 8000 and

0.5 M of NaCl followed by incubation at 4°C for 16 hrs. The phage was recovered by centrifugation at 4,000 rpm (3,500xg) for 20 min at 4°C on a GS-6R table top centrifuge (Beckman). The pellet was resuspended with 2 ml of phage buffer (1 mM MgSO₄, 5 mM MgCl₂, 80 mM NaCl and 0.1% Gelatin). The phage suspension was
5 extracted with 1 volume of chloroform and further purified by centrifugation on a cesium chloride step gradient as described in Sambrook et al. (1989), using a TLS 55 rotor centrifuged in an Optima TLX ultracentrifuge (Beckman) for 2 h at 28,000 rpm (67,000xg) at 4°C. Banded phage was collected and ultracentrifuged again on an isopycnic cesium chloride gradient (1.45 g/ml) at 40,000 rpm (64,000xg) for 24 h at
10 4°C using a TLV rotor (Beckman). The phage was harvested and dialyzed for 4 h at room temperature against 4 L of dialysis buffer consisting of 10 mM NaCl, 50 mM Tris-HCl [pH 8] and 10 mM MgCl₂. Phage DNA was prepared from the phage suspension by adding 20 mM EDTA, 50 mg/ml Proteinase K and 0.5% SDS and incubating for 1 h at 65°C, followed by successive extractions with 1 volume of
15 phenol, 1 volume of phenol-chloroform and 1 volume of chloroform. The DNA was then dialyzed overnight at 4°C against 4 L of TE (10 mM Tris pH 8.0, 1mM EDTA).

Example 2. DNA sequencing of Bacteriophage 77 genome

Four micrograms of phage 77 DNA was diluted in 200 µl of TE (10 mM Tris, [pH 8.0], 1 mM EDTA) in a 1.5 ml eppendorf tube and sonication was performed
20 (550 Sonic Dismembrator™, Fisher Scientific). Samples were sonicated under an amplitude of 3 µm with bursts of 5 s spaced by 15 s cooling in ice/water for 3 to 4 cycles. The sonicated DNA was then size fractionated by electrophoresis on 1% agarose gels utilizing TAE (1 x TAE is: 40 mM Tris-acetate, 1 mM EDTA [pH 8.0])
25 as the running buffer. Fractions ranging from 1 to 2 kbp were excised from the agarose gel and purified using a commercial DNA extraction system according to the instructions of the manufacturer (Qiagen), with a final elution of 50 µl of 1 mM Tris (pH 8.5).

The ends of the sonicated DNA fragments were repaired with a combination of
30 T4 DNA polymerase and the Klenow fragment of E. coli DNA polymerase I, as follows. Reactions were performed in a reaction mixture (final volume, 100 µl) containing sonicated phage DNA, 10 mM Tris-HCl [pH 8.0], 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 50 µg/ml BSA, 100 µM of each dNTP and 15 units of T4 DNA polymerase (New England Biolabs) for 20 min at 12°C followed by addition of 12.5
35 units of Klenow large fragment (New England Biolabs) for 15 min at room - temperature. The reaction was stopped by two phenol/chloroform extractions and the

DNA was precipitated with ethanol and the final DNA pellet was resuspended in 20 μ l of H₂O.

Blunt-ended DNA fragments were cloned by ligation directly into the *Hinc* II site of pKSII+ vector (New England Biolabs) dephosphorylated by treatment with calf intestinal alkaline phosphatase (New England Biolabs)-treated pKS II+ vector
5 (Stratagene). A typical ligation reaction contained 100 ng of vector DNA, 2 to 5 μ l of repaired sonicated phage DNA (50-100 ng) in a final volume of 20 μ l containing 800 units of T4 DNA ligase (New England Biolabs) and was incubated overnight at 16°C. Transformation and selection of bacterial clones containing recombinant plasmids was
10 performed in *E. coli* DH10 β according to standard procedures (Sambrook et al., 1989).

Recombinant clones were picked from agar plates into 96-well plates containing 100 μ l LB and 100 μ g/ml ampicillin and incubated at 37°C. The presence of phage DNA insert was confirmed by PCR amplification using T3 and T7 primers
15 flanking the *Hinc* II cloning site of the pKS II+ vector. PCR amplification of foreign insert was performed in a 15 μ l reaction volume containing 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.02% gelatin, 1 μ M primer, 187.5 μ M each dNTP, and 0.75 units *Taq* polymerase (BRL). The thermocycling parameters were as follows: 2 min initial denaturation at 94°C for 2 min, followed by 20 cycles of 30 sec
20 denaturation at 94°C, 30 sec annealing at 57°C, and 2 min extension at 72°C, followed by a single extension step at 72°C for 10 min. Clones with insert sizes of 1 to 2 kbp were selected and plasmid DNA was prepared from the selected clones using QIAprep™ spin miniprep kit (Qiagen).

The nucleotide sequence of the extremities of each recombinant clone was
25 determined using an ABI 377-36 automated sequencer with two types of chemistry: ABI prism Big Dye™ primer or ABI prism Big Dye™ terminator cycle sequencing ready reaction kit (Applied Biosystems). To ensure co-linearity of the sequence data and the genome, all regions of phage genome were sequenced at least once from both directions on two separate clones. In areas that this criteria was not initially met, a
30 sequencing primer was selected and phage DNA was used directly as sequencing template employing ABI prism Big Dye™ terminator cycle sequencing ready reaction kit.

Example 3. Bioinformatic management of primary nucleotide sequence from
35 Phage 77.

Phage 77 sequence contigs were assembled using Sequencher™ 3.1 software (GeneCodes). To close contig gaps, sequencing primers were selected near the edge of

the contigs. Phage DNA was used directly as sequencing template employing ABI prism BIG DYE™ terminator cycle sequencing ready reaction kit. The complete sequence of bacteriophage 77 is shown in Table 2.

A software program was developed and used on the assembled sequence of bacteriophage 77 to identify all putative ORFs larger than 33 codons. Other ORF identification software can also be utilized, preferably programs which allow alternative start codons. The software scans the primary nucleotide sequence starting at nucleotide #1 for an appropriate start codon. Three possible selections can be made for defining the nature of the start codon; I) selection of ATG, II) selection of ATG or GTG, and III) selection of either ATG, GTG, TTG, CTG, ATT, ATC, and ATA. This latter initiation codon set corresponds to the one reported by the NCBI (<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c>) for the bacterial genetic code.

When an appropriate start codon is encountered, a counting mechanism is employed to count the number of codons (groups of three nucleotides) between this start codon and the next stop codon downstream of it. If a threshold value of 33 is reached, or exceeded, then the sequence encompassed by these two codons (start and stop codons) is defined as an ORF. This procedure is repeated, each time starting at the next nucleotide following the previous stop codon found, in order to identify all the other putative ORFs. The scan is performed on all three reading frames of both DNA strands of the phage sequence.

Sequence homology (BLAST) searches for each ORF are then carried out using an implementation of BLAST programs, although any of a variety of different sequence comparison and matching programs can be utilized as known to those skilled in the art. Downloaded public databases used for sequence analysis include:

- i) non-redundant GenBank (<ftp://ncbi.nlm.nih.gov/blast/db/nr.Z>),
- ii) Swissprot (<ftp://ncbi.nlm.nih.gov/blast/db/swissprot.Z>);
- iii) vector (<ftp://ncbi.nlm.nih.gov/blast/db/vector.Z>);
- iv) pdbaa databases (<ftp://ncbi.nlm.nih.gov/blast/db/pdbaa.Z>);
- v) *S. aureus* NCTC 8325 (<ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa>);
- vi) *streptococcus pyogenes* (<ftp://ftp.genome.ou.edu/pub/strep/strep-1k.fa>);
- vii) *Streptococcus pneumoniae* (ftp://ftp.tigr.org/pub/data/s_pneumoniae/gsp.contigs.112197.Z);
- viii) *Mycobacterium tuberculosis* CSU#9 (ftp://ftp.tigr.org/pub/data/m_tuberculosis/TB_091097.Z) and
- ix) *pseudomonas aeruginosa* (<http://www.genome.washington.edu/pseudo/data.html>).

The results of the homology searches performed on the ORFs is shown in Table 5.

Example 4. Subcloning of Bacteriophage 77 ORFs into a Staph A inducible expression system.

The shuttle vector pT0021, in which the firefly luciferase (*lucFF*) expression is controlled by the *ars* (arsenite) promoter/operator (Tauriainen et al., 1997), was modified in the following fashion. Two oligonucleotides corresponding to a short antigenic peptide derived from the hemagglutinin protein of influenza virus (HA epitope tag) were synthesized (Field et al., 1988). The sense strand HA tag sequence (with *Bam*HI, *Sal*I and *Hind*III cloning sites) is:

5'-gatcccggtcgaccaagctTACCCATACGACGTCCCAGACTACGCCAGCTGA-3' (where upper case letters denote the nucleotide sequence of the HA tag); the antisense strand HA tag sequence (with a *Hind*III cloning site) is:

5'-agctTCAGCTGGCGTAGTCTGGGACGTCGTATGGGTAAagcttggtcgaccgg-3' (where upper case letters denote the sequence of the HA tag). The two HA tag oligonucleotides were annealed and ligated into pT0021 vector which had been digested with *Bam*HI and *Hind*III. This manipulation resulted in replacement of the *lucFF* gene by the HA tag. This modified shuttle vector containing the *arsenite* inducible promoter, the *arsR* gene, and HA tag was named pTHA. A diagram outlining our modification of pT0021 to generate pTHA is shown in Fig. 1A.

Each ORF, encoded by Bacteriophage 77, larger than 33 amino acids and having a Shine-Dalgarno sequence upstream of the initiation codon was selected for functional analysis for bacterial inhibition. In total, 98 ORFs were selected and screened as detailed below. A list of these is presented in Table 3. Each individual ORF, from initiation codon to last codon (excluding the stop codon), was amplified from phage genomic DNA using the polymerase chain reaction (PCR). For PCR amplification of ORFs, each sense strand primer targets the initiation codon and is preceded by a *Bam*HI restriction site (5'-cgggatcc-3') and each antisense oligonucleotide targets the penultimate codon (the one before the stop codon) of the ORF and is preceded by a *Sal*I restriction site (5'-gcgtcgaccg-3'). The PCR product of each ORF was gel purified and digested with *Bam*HI and *Sal*I. The digested PCR product was then gel purified using the Qiagen kit as described, ligated into *Bam*HI and *Sal*I digested pTHA vector, and used to transform *E. coli* bacterial strain DH10 β (as described above). As a result of this manipulation, the HA tag is set inframe with the ORF and is positioned at the carboxy terminus of each ORF (pTHA/ORF clones). Recombinant pTHA/ORF clones were picked and their insert sizes were confirmed by PCR analysis

using primers flanking the cloning site. The names and sequences of the primers that were used for the PCR amplification were: HAF:

'TATTATCCAAACTTGAACA'; HAR: 'CGGTGGTATATCCAGTGATT'. The

- 5 primers HAF and HAR. In cases where verification of ORF sequence could not be achieved by one pass with the sequencing primers, additional internal primers were selected and used for sequencing.

- 10 *Staphylococcus aureus* strain RN4220 (Kreiwirth et al., 1983) was used as a recipient for the expression of recombinant plasmids. Electoporation was performed essentially as previously described (Schenk and Laddaga, 1992). Selection of recombinant clones was performed on Luria-Broth agar (LB-agar) plates containing 30 µg/ml of kanamycin.

- For each ORF introduced in the pTHA plasmid, 3 independent transformants were isolated and used to individually inoculate cultures in 5 ml of TSB containing 30µg/ml kanamycin, followed by growth to saturation (16 hrs at 30°C). An aliquot of this stationary phase culture was used to generate a frozen glycerol stock of the transformant (stored at - 80°C). The remaining culture was used for plasmid DNA extraction. Bacterial cells were harvested by centrifugation at 3000 x g at 22°C for 5 min. The pellet was resuspended in 200 µl 25% sucrose containing 25U/ml of 20 lysostaphin and incubated for 15 min at 37°C. Then, 400µl of alkaline SDS solution (3% SDS, 0.2N NaOH) were added, well mixed and incubated for 7 min at room temperature. After the alkaline SDS treatment, 300µl of ice-cold 3M sodium acetate pH 4.8 were added, and the mix is immediately spun at 13000g for 15 min at room temperature. The supernatant was transferred to a new 1.5 ml conical centrifuge tube 25 and 650µl of isopropanol (stored at room temperature) were added. The mix was then centrifuged at 13,000 x g for 5 min. The supernatant fluid was discarded, the pellet washed with 70% ethanol, and resuspended in 320 µl sterile distilled water.

- The presence of individual phage 77 ORF DNA inserts in the plasmid was verified by PCR amplification using 1.5 µl transformant miniprep DNA in a PCR 30 with primers flanking the cloning site of ORF in pTHA vector (HAF and HAR). The composition of the PCR reaction and the cycling parameters are identical to those employed for library screening described above.

- Example 5. Functional assay for bacterial inhibitory activity of bacteriophage 77**
35 **ORFs.**

The anti-microbial activity of individual phage 77 ORFs was monitored by two growth inhibitory assays, one on solid agar medium, the other in liquid medium.

In general, *Staphylococcus* bacteria transformed with expression plasmids containing individual ORFs were grown in normal TSA medium and stored in 19% glycerol. At pre-determined times, arsenite was added to the culture to induce transcription of the phage 77 ORFs cloned immediately downstream from an arsenite-inducible promoter in the pTHA expression plasmid.

The effect of ORF induction on bacterial growth characteristics was then monitored and quantitated. The growth inhibition assay on solid medium was performed by streaking pTHA/ORF containing *S. aureus* transformant onto LB-Kn and TSA-Kn plates containing increasing concentrations of sodium arsenite (0; 2.5; 5; and 7.5 μ M). Arsenite is used to induce the expression of cloned DNA in pTHA vector. In parallel, 3 μ l of 1/10 and 1/100 dilutions of the frozen cultures of the pTHA/ORF transformants were spotted as single drops onto LB-Kn and TSA-Kn plates containing increasing concentration of sodium arsenite (0; 2.5; 5; and 7.5 μ M). The plates were then incubated 16 hrs at 37°C, and the effect of arsenite-induced ORF expression on bacterial growth was monitored and quantitated by comparing the extent to that seen in control plates. As positive controls for growth inhibition, the *holin/lysine* genes of the *Staphylococcus aureus* phage Twort (Loessner et al., 1998) was subcloned into the pTHA *ars* inducible vector and used.

For the growth inhibition assay in liquid medium, stationary phase cultures were prepared by inoculating 2.5ml TSB-Kn with frozen *S. aureus* RN4220 transformants containing phage 77 ORFs cloned in pTHA vector followed by incubation for 16 hrs at 37°C. These cultures were then diluted 1/100 in the same medium, and the bacteria were allowed to grow for 2 hrs at 37°C to reach early log phase. 150 μ l of such culture were then mixed with 2.35 ml TSB-Kn medium with or without arsenite (the final concentration of arsenite in the medium was 0 or 5 μ M arsenite). After 3.5 hrs incubation at 37°C with shaking at 250 rpm, 100 μ l of bacterial culture was removed from each tube for OD₅₆₅ measurement. Serial ten-fold dilutions of the culture in buffered saline solution (0.85% NaCl) were then spotted onto TSB-Kn plates. The plates were incubated at 37°C 16 hrs and the number of surviving colonies counted the following day. The growth inhibitory property of individual ORFs was then quantitated by comparing CFU numbers under normal or arsenite-induction conditions. A schematic flow of the inhibition analysis is shown in Fig. 3 (also applicable to inhibition analysis for the other phage and bacteria pointed out herein). Inhibition results are shown in Figures 4A-C.

Example 6: Identification of Cecropin Signature Motif in *Staphylococcus aureus* Bacteriophage 3A ORF

The genome for *S. aureus* bacteriophage 3A was determined and the sequence was analyzed essentially as described for bacteriophage 77 in the examples above. Upon blast analysis of the identified open reading frames of phage 3A, the presence of an amino acid sequence corresponding to a cecropin signature motif was observed.

- 5 This motif (WDGHKTLEK) is located at position aa 481-489. Cecropins were originally identified in proteins from the cecropia moth and are recognized as potent antibacterial proteins that constitute an important part of the cell-free immunity of insects. Cecropins are small proteins (31-39 amino acid residues) that are active against both Gram-positive and Gram-negative bacteria by disrupting the bacterial
10 membranes. Although the mechanisms by which the cecropins cause cell death are not fully understood, it is generally thought to involve channel formation and membrane destabilization.

- The identification of a motif corresponding to a known inhibitor suggests that the product of ORF002 is also an inhibitory compound. Such inhibitory activity can
15 be confirmed as described herein or by other methods known in the art. Confirmation of the inhibitory activity would indicate that the ORF product could serve as the basis for construction of mimetic compounds and other inhibitors directed to the target of the ORF002 product.

Boman & Hultmark, 1987, *Ann. Rev. Microbiol.* 41:103-126.

- 20 Boman, 1991, *Cell* 65:205-207.

Boman et al., 1991, *Eur. J. Biochem.* 201:23-31.

Wang et al., *J. Biol. Chem.* 273:27438-27448.

Example 7. Growth of *Staphylococcus aureus* bacteriophage 44AHJD:

- 25 *Staphylococcus aureus* propagating strain (PS 44A) (Felix d'Herelle Reference Centre #HER 1101) was used as a host to propagate its respective phage 44AHJD (Felix d'Herelle Reference Centre #HER 101). Two rounds of plaque purification of phage 44AHJD were performed on soft agar essentially as described in Sambrook *et al.* (1989). Briefly, the *Staphylococcus aureus* PS strain was grown overnight at 37°C
30 in Nutrient Broth [NB: 3 g Bacto Beef Extract, 5 g Bacto peptone per liter, (Difco Laboratories # 0003-17-8), supplemented with 0.5% NaCl]. The culture was then diluted 20 fold in NB and incubated at 37°C until an OD₅₄₀ of 0.2. In order to obtain single plaques, phage 44AHJD was subjected to 10-fold serial dilutions using the phage buffer (1 mM MgSO₄, 5 mM MgCl₂, 80 mM NaCl and 0.1% Gelatin) and 10 µl
35 were used to infect 0.5 ml of the cell suspension in the presence of 400 µg/ml of

CaCl₂. After incubation of 15 min at room temperature, 2 ml of melted soft agar (NB supplemented with 0.6% of agar) were added to the mixture and poured onto the surface of 100 mm nutrient agar plates (3 g Bacto Beef extract, 5 g Bactopeptone, 0.5% NaCl and 15 g of Bacto agar per liter (Difco Laboratories # 0001-17-0). After
5 overnight incubation at 37°C, a single plaque was isolated, resuspended in 1ml of phage buffer by end over end rotation for 2 h at room temperature and the phage suspension was diluted and used for a second infection as described above. After overnight incubation at 37°C, a single plaque was isolated and used as a stock.

Large scale purification of bacteriophage and preparation of phage DNA was
10 as follows.

The propagation method was carried out by using the agar layer method described by Swanstörn and Adams (1951). Briefly, the PS 44A strain was grown to stationary phase overnight at 37°C in Nutrient Broth. The culture was then diluted 20x in NB and incubated at 37°C until the A₅₄₀ = 0.2. The suspension (15x10⁷ Bacteria)
15 was then mixed with 15x10⁵ phage particles to give a ratio of 100-bacteria/phage particle in the presence of 400 µg/ml of CaCl₂. After incubation of 15 min at room temperature, 7.5 ml of melted soft agar were added to the mixture and poured onto the surface of 150 mm nutrient agar plates and incubated overnight at 37°C. To collect the lysate, 20 ml of NB were added to each plate and the soft agar layer was collected by
20 scrapping off with a clean microscope slide and shaken vigorously for 5 min to break up the agar. The mixture was then centrifuged for 10 min at 4,000 rpm (2,830 xg) using a JA-10 rotor (Beckman) and the supernatant (lysate) is collected and subjected to a treatment with 10 µg/ml of DNase I and RNase A for 30 min at 37°C. To precipitate the phage particles, 10% (w/v) of PEG 8000 and 0.5 M of NaCl were
25 added to the lysate and the mixture was incubated on ice for 16 h. The phage was recovered by centrifugation at 4,000 rpm (3,500 xg) for 20 min at 4°C on a GS-6R table top centrifuge (Beckman).

The pellet was resuspended with 2 ml of phage buffer (1 mM MgSO₄, 5 mM MgCl₂, 80 mM NaCl and 0.1% Gelatin). The phage suspension was extracted with 1
30 volume of chloroform and further purified by centrifugation on a preformed cesium chloride step gradient as described in Sambrook *et al.* (1989), using a TLS 55 rotor and centrifuged in an Optima TLX ultracentrifuge (Beckman) for 2 h at 28,000 rpm (67,000 xg) at 4°C. Banded phage was collected and ultracentrifuged again on an

isopycnic cesium chloride gradient (1.45 g/ml) at 40,000 rpm (64,000 x g) for 24 h at 4°C using a TLV rotor (Beckman). The phage was harvested and dialyzed for 4 h at room temperature against 4 L of dialysis buffer consisting of 10 mM NaCl, 50 mM Tris-HCl [pH 8] and 10 mM MgCl₂. Phage DNA was prepared from the phage suspension by adding 20 mM EDTA, 50 µg/ml Proteinase K and 0.5% SDS and incubating for 1 h at 65°C, followed by successive extractions with 1 volume of phenol, 1 volume of phenol-chloroform and 1 volume of chloroform. The DNA was then dialyzed overnight at 4°C against 4 L of TE (10 mM Tris-HCl [pH 8.0], 1mM EDTA).

10

Example 8. DNA sequencing of the Bacteriophage 44 AHJD genome.

Four mg of phage DNA was diluted in 200 µl of TE pH 8.0 in a 1.5 ml eppendorf tube and sonication was performed (550 Sonic Dismembrator, Fisher Scientific). Samples were sonicated under an amplitude of 3 µm with bursts of 5 s spaced by 15 s cooling in ice/water for 3 to 4 cycles and size fractionated on 1% agarose gels. The sonicated DNA was then size fractionated by gel electrophoresis. Fractions ranging from 1 to 2 kbp were excised from the agarose gel and purified using a commercial DNA extraction system according to the instructions of the manufacturer (Qiagen) and eluted in 50 µl of 1mM Tris-HCl [pH 8.5].

20

The ends of the sonicated DNA fragments were repaired with a combination of T4 DNA polymerase and the Klenow fragment of *E. coli* DNA polymerase 1 as follows. Reactions were performed in a final volume of 100 µl containing DNA, 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 5 µg BSA, 100 µM of each dNTP and 15 units of T4 DNA polymerase (New England Biolabs) for 20 min at 12°C followed by addition of 12.5 units of Klenow fragment (New England Biolabs) for 15 min at room temperature. The reaction was stopped by two phenol/chloroform extractions and the DNA was ethanol precipitated and resuspended in 20 µl of H₂O.

25

Cloning of the sonicated phage DNA into pKSII vector and transformation:

30

Blunt-ended DNA fragments were cloned by ligation directly into the *HincII* site of the pKSII vector (Stratagene) dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs). A typical reaction contained 100 ng of vector, 2

to 5 µl of repaired sonicated phage DNA (50-100 ng) in a final volume of 20 µl containing 800 units of T4 DNA ligase (New England Biolabs) overnight at 16°C. Transformation and selection of positive clones was performed in the host strain DH10 β of *E. coli* using ampicillin as a selective antibiotic as described in Sambrook *et al.* (1989).

Recombinant clones were picked from agar plates into 96-well plates containing 100 µl LB and 100 µg/ml ampicillin and incubated at 37°C. The presence of phage DNA insert was confirmed by PCR amplification using T3 and T7 primers flanking the *HincII* cloning site of the pKS vector. PCR amplification of the potential foreign inserts was performed in a 15 µl reaction volume containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.02% gelatin, 1 mM primer, 187.5 µM each dNTP, and 0.75 units *Taq* polymerase (BRL). The thermocycling parameters were as follows: 2 min initial denaturation at 94°C for 2 min, followed by 20 cycles of 30 sec denaturation at 94°C, 30 sec annealing at 58°C, and 2 min extension at 72°C, followed by a single extension step at 72°C for 10 min. Clones with insert sizes of 1 to 2 kbp were selected and plasmid DNA was prepared from the selected clones using the QIAprep™ spin miniprep kit (Qiagen).

The nucleotide sequence of the extremities of each recombinant clone was determined using an ABI 377-36 automated sequencer with two types of chemistry: ABI prism BigDye™ primer cycle sequencing (21M13 primer: #403055)(M13REV primer: #403056) or ABI prism BigDye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). To ensure co-linearity of the sequence data and the genome, all regions of the phage genome were sequenced at least once from both directions on two separate clones. In areas that this criteria was not initially met, a sequencing primer was selected and phage DNA was used directly as sequencing template employing ABI prism BigDye™ terminator cycle sequencing ready reaction kit.

Example 9. Bioinformatic management of primary nucleotide sequence.

Sequence contigs were assembled using Sequencher™ 3.1 software (GeneCodes). To close contig gaps, sequencing primers were selected near the edge of the contigs. Phage DNA was used directly as sequencing template employing ABI

prism BigDye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). The complete sequence of *Staphylococcus aureus* bacteriophage 44AHJD is shown in Table 16.

A software program was used on the assembled sequence of bacteriophage 44AHJD to identify all putative ORFs larger than 33 codons. The software scans the primary nucleotide sequence starting at nucleotide #1 for an appropriate start codon. Three possible selections can be made for defining the nature of the start codon; I) selection of ATG, II) selection of ATG or GTG, and III) selection of either ATG, GTG, TTG, CTG, ATT, ATC, and ATA. This latter initiation codon set corresponds to the one reported by the NCBI(<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c>) for the bacterial genetic code. When an appropriate start codon is encountered, a counting mechanism is employed to count the number of codons (groups of three nucleotides) between this start codon and the next stop codon downstream of it. If a threshold value of 33 is reached, or exceeded, then the sequence encompassed by these two codons is defined as an ORF. This procedure is repeated, each time starting at the next nucleotide following the previous stop codon found, in order to identify all the other putative ORFs. The scan is performed on all three reading frames of both DNA strands of the phage sequence. The predicted ORFs for bacteriophage 44AHJD are listed in Tables 17 & 18.

Sequence homology searches for each ORF were carried out using an implementation of blast programs. Downloaded public databases used for sequence analysis include:

- (i) non-redundant GenBank (<ftp://ncbi.nlm.nih.gov/blast/db/nr.Z>),
- ii) Swissprot (<ftp://ncbi.nlm.nih.gov/blast/db/swissprot.Z>);
- iii) vector (<ftp://ncbi.nlm.nih.gov/blast/db/vector.Z>);
- iv) pdbaa databases (<ftp://ncbi.nlm.nih.gov/blast/db/pdbaa.Z>);
- v) *Staphylococcus aureus* NCTC 8325 (<ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa>);
- vi) *Staphylococcus pyogenes* (ftp://ftp.tigr.org/pub/data/s_pneumoniae/gsp.contigs.112197.Z);
- vii) PRODOM (ftp://ftp.toulouse.inra.fr/pub/prodom/current_release/prodom99_1.förblast.gz);
- viii) DOMO (<ftp://ftp.infobiogen.fr/pub/db/domo/>);

ix) TREMBL (ftp://www.expasy.ch/databases/sp_tr_nrd/fasta/)

The results of the homology searches performed on the ORFs of bacteriophage 44AHJD are shown in Tables 19 & 20.

5 Example 10. Sub-Cloning of Bacteriophage 44 AHJD ORFs.

Expression preferably utilizes a shuttle expression vector which is arranged such that expression of the exogenous bacteriophage 44 AHJD ORF sequence is inducible. For example, the shuttle vector pT0021, in which the firefly luciferase (*lucFF*) expression is controlled by the *ars* (arsenite) promoter/operator (Tauriainen et al., 1997), can be modified in the following fashion. Two oligonucleotides corresponding to a short antigenic peptide derived from the hemagglutinin protein of influenza virus (HA epitope tag) were synthesized (Field et al., 1988). The sense strand HA tag sequence (with *Bam*HI, *Sal*I and *Hind*III cloning sites) is:

5'-gatcccggtcgaccaagctTACCCATACGACGTCCCAGACTACGCCAGCTGA-3'

15 (where upper case letters denote the nucleotide sequence of the HA tag); the antisense strand HA tag sequence (with a *Hind*III cloning site) is:

5'-agctTCAGCTGGCGTAGTCTGGGACGTCGTATGGGTAaagcttggtcgaccgg-3'

(where upper case letters denote the sequence of the HA tag). The two HA tag oligonucleotides were annealed and ligated into pT0021 vector which had been digested with *Bam*HI and *Hind*III. This manipulation resulted in replacement of the *lucFF* gene by the HA tag. This modified shuttle vector containing the *arsenite* inducible promoter, the *arsR* gene, and HA tag was named pTHA. A diagram outlining our modification of pT0021 to generate pTHA is shown in Fig. 1A (another useful vector construct is shown in Fig. 1B).

25 Each ORF, encoded by Bacteriophage 44 AHJD, larger than 33 amino acids and having a Shine-Dalgarno sequence upstream of the initiation codon can be selected for functional analysis for bacterial inhibition. Each individual ORF, from initiation codon to last codon (excluding the stop codon), can be amplified from phage genomic DNA using the polymerase chain reaction (PCR). For PCR amplification of ORFs, each sense strand primer targets the initiation codon and is preceded by a *Bam*HI restriction site (5'-cggtatcc-3') and each antisense oligonucleotide targets the pentultimate codon (the one before the stop codon) of the ORF and is preceded by a *Sal*I restriction site (5'-ggtgacg-3'). The PCR product of each ORF can be gel

purified and digested with *Bam*HI and *Sal*I. The digested PCR product can then be gel purified using the Qiagen kit as described, ligated into *Bam*HI and *Sal*I digested pTHA vector, and used to transform *E. coli* bacterial strain DH10 β (as described above). As a result of this manipulation, the HA tag is set inframe with the ORF and is positioned at the carboxy terminus of each ORF (pTHA/ORF clones). Recombinant pTHA/ORF clones will be picked and their insert sizes were confirmed by PCR analysis using primers flanking the cloning site. The following primers can be used for PCR amplification: HAF: 'TATTATCCAAACTTGAACA'; HAR: 'CGGTGGTATATCCAGTGATT'. The sequence integrity of cloned ORFs can be verified directly by DNA sequencing using primers HAF and HAR. In cases where verification of ORF sequence can not be achieved by one pass with the sequencing primers, additional internal primers will be selected and used for sequencing.

Staphylococcus aureus strain RN4220 (Kreiwirth et al., 1983) will be used as a recipient for the expression of recombinant plasmids. Electoporation will be performed essentially as previously described (Schenk and Laddaga, 1992). Selection of recombinant clones will be performed on Luria-Broth agar (LB-agar) plates containing 30 μ g/ml of kanamycin.

Alternatively, a constitutive promoter can be used to drive expression of the introduced ORF, and compare cell growth to control bacterial cells containing the parental vector lacking any introduced phage ORF. Recombinant plasmids will be introduced into *Staphylococcus aureus* strain RN4220 (Kreiwirth et al., 1983) using electoporation as previously described (Schenk and Laddaga, 1992).

Cloning of ORFs with a Shine-Dalgarno sequence

ORFs with a Shine-Dalgarno sequence are selected for functional analysis of bacterial killing. Each ORF, from initiation codon to last codon (excluding the stop codon), can be amplified by PCR from phage genomic DNA. For PCR amplification of ORFs, each sense strand primer starts at the initiation codon and is preceded by a restriction site and each antisense strand starts at the last codon (excluding the stop codon) and is preceded by a different restriction site. The PCR product of each ORF will be gel purified and digested with the restriction enzymes with sites contained on the PCR oligonucleotides. The digested PCR product is then gel purified using the Qiagen kit, ligated into the modified shuttle vector, and used to transform bacterial strain DH10. Recombinant clones are then picked and their insert sizes confirmed by

PCR analysis using primers flanking the cloning site as well as restriction digestion. The sequence fidelity of cloned ORFs can be verified by DNA sequencing using the same primers as used for PCR. In the cases that the verification of ORFs can not be achieved by one path of sequencing using primers flanking the cloning site internal
5 primers can be selected and used for sequencing. Recombinant plasmids can be introduced into *Staphylococcus aureus* strain RN4220 (Kreiwirth et al., 1983) using electroporation as previously described (Schenk and Laddaga, 1992).

Induction of gene expression from the *ars* promoter.

If an inducible promoter is used, e.g., the *ars* promoter, induction can be
10 assessed, for example, in either of the two methods.

1. Screening on agar plates

The functional identification of killer ORFs can be performed by spreading an aliquot of *S. aureus* transformed cells containing phage 44 AHJD ORFs onto agar plates containing different concentrations of sodium arsenite (0; 2.5; 5; and 7.5 μ M). The
15 plates are incubated overnight at 37°C, after which a growth inhibition of the ORF transformants on plates that contain arsenite are compared to plates without arsenite.

2. Quantification of growth inhibition in liquid medium

Cells containing different recombinant plasmids can be grown for overnight at 37°C in LB medium supplemented with the appropriate antibiotic selection. These are
20 then diluted to the mid log phase ($OD_{540}=0.2$) with fresh media containing antibiotic and transferred to 96-well microtitration plates (100 μ l/well). Inducer is then added at different final concentrations (ranging from 2.5 to 10 μ M) and the culture incubated for an additional 2 hrs at 37°C. The effect of expression of the phage 44 AHJD ORFs on bacterial cell growth is then monitored by measuring the OD_{540} and comparing the
25 rate of growth to the culture not containing inducer. [As positive controls for growth inhibition, the *kilA* gene of phage lambda (Reisinger, GR., Rietsch, A., Lubitz, W. and Blasi, U. 1993 *Virology* #193: 1033-1036), and the *holin/lysin* genes of the *Staphylococcus aureus* phage Twort (Loessner, MJ., Gaeng, S., Wendlinger, G., Maier, SK. and Scherer, S. 1998. *FEMS Microbiology Letters* #162:265-274) can be
30 subcloned into the *ars* inducible vector. An aliquot of the induced and uninduced culture can also be plated out on agar plates containing an appropriate antibiotic selection but lacking inducer. Following incubation overnight at 37°C, the number of

colonies is counted. Any ORF showing bacteriostatic activity will show a lower, but detectable, number of colonies on the agar plates when grown in the presence of inducer as compared to when grown in the absence of inducer. Any ORF showing full bacteriocidal activity will show no colonies on the agar plates, when grown in the presence of inducer as compared to when grown in the absence of inducer.

REFERENCES

- Ackermann, H-W. and DuBow, M. S. (1987). Viruses of Prokaryotes. Volumes I and II. CRC Press, Boca Raton, Florida.
- Tenover, F.C. and McGowan Jr., J.E. (1998). Bacterial Infections of Humans. Epidemiology and Control.(A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 83-93.
- Rusterholtz, K., and Pohlschroder, M. (1999). Cell 96, 469-470.
- Gray, B.M. (1998). Bacterial Infections of Humans. Epidemiology and Control.(A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 673-711.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, New York. Cold Spring Harbor Laboratory Press.
- Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J.
- Rost B, and Sander C. (1996). Ann. Rev. Biophys. Biomol. Struct. 25, 113-136.
- Martin, A.C., Lopez, R., Garcia, P. (1998). J Bacteriol 180, 210-217.
- Steiner, M., Lubitz, W., Blasi, U. (1993). J. Bacteriol. 175, 1038-1042.
- Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., and Elledge, S.J. (1993). Genes Dev. 7, 555-569.
- Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998). Mol Cell Biol. 18, 2697-2711.

- Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura, Y., and Miki, Y. (1998). *Genes, Chromosomes & Cancer* 21, 217-222.
- 5 Endo, T.A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. (1997). *Nature* 387, 921-924.
- 10 Karimova, G., Pidoux, J., Ullmann, A., Ladant, D. (1998) *Proc. Natl. Acad. Sci.* 95, 5752-5756.
- Sopta, M., Carthew, R.W., and Greenblatt, J. (1995) *J. Biol. Chem.* 260, 10353-10369.
- 15 Qin, J., Fenyo, D., Zhao, Y., Hall, W.W., Chao, D.M., Wilson, C.J., Young, R.A. and Chait, B.T. (1997). *Anal. Chem.* 69, 3995-4001.
- Swanström, M. and Adams, M.H. (1951). *Proc. Soc. Exptl. Biol. & Med.* 78: 372-375.
- 20 Røder, B.L., Wandall, D. A., Frimødt-Møller, N., Epersen, F., Skinhøj, P. and Rosdahl, T. (1999). *Arch. Intern. Med.* 159: 462-469.
- 25 Sanabria, T.J., Albert, J.S., Goldberg, R., Pape, L.A. and Cheeseman, S.H. (1990). *Arch. Intern. Med.* 150: 1305-1309.
- Frimødt-Møller, N., Epersen, F., Skinhøj, P. and Rosdahl, V.T. (1997). *Clin. Microbiol. Infect.* 3: 297-305.
- 30 Harbath, S., Rutschmann, O., Sudre, P. and Pittet, D. (1998). *Arch. Intern. Med.* 158: 182-189.
- Steinberg, J.P., Clark, C.C. and Hackman, B.O. (1996). *Clin. Infect. Dis.* 23: 255-259.
- 35 Field, J., Nikawa, J.-I., Broek, D., MacDonald, B., Rodgers, L., Wilson, I.A., Lemer, R.A., and Wigler, M. (1988). Purification of a RAS-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol. Cell. Biol.* 8: 2159-2165.
- 40 Kreiswirth, BN., Lofdahl, S., Belley, MJ., O'Reilly, M., Shlievert, PM., Bergdoll, MS. and Novicks, RP. (1983) *Nature* 305: 709-712.

Schenk, S. and Laddaga, RA. (1992) FEMS Microbiology Letters 94: 133-138.

Cohen, M.L. (1992) Science 257, 1050-1055.

5

Example 11. Growth of *Enterococcus* bacteriophage 182 and purification of genomic DNA.

The *Enterococcus* propagating strain (PS) (*Enterococcus* sp. Group D, Felix d'Herelle Reference Centre #HER 1080) was used as host to propagate its respective
10 phage 182 (Felix d'Herelle Reference Centre #HER 80). Two rounds of plaque purification of phage 182 were performed on soft agar essentially as described in Sambrook *et al.* (1989). Briefly, the *Enterococcus* sp. PS strain was grown overnight at 37°C in Tryptic Soy Broth [TSB: 17 g Bacto tryptone, 3 g Bacto soytone, 2.5 g Bacto dextrose, 5 g Sodium chloride, and 2.5 g Dipotassium phosphate per liter
15 (Difco Laboratories (#0370-17-3))]. The culture was then diluted 20 fold in TSB and incubated at 37°C until the $OD_{540} = 0.2$ (early log phase) with constant agitation. In order to obtain single plaques, phage 182 was subjected to 10 fold serial dilutions using the phage buffer (1 mM $MgSO_4$, 5 mM $MgCl_2$, 80 mM NaCl and 0.1% Gelatin (w/v)) and 10 l of each dilution was used to infect 0.5 ml of the bacterial cell
20 suspension. After incubation at 15 min at 37°C, 2 ml of melted soft agar (TSB supplemented with 0.6% agar) was added to the mixture and poured onto the surface of 100 mm Tryptic Soy Agar plates [TSA: 15 g Tryptone peptone, 5 g Soytone peptone, 5 g Sodium chloride and 15 g of Agar per liter (Difco Laboratories #0369-17)]. After overnight incubation at 37°C, a single plaque was isolated, resuspended in
25 1 ml of phage buffer by end over end rotation for 2 hrs at room temperature, and the phage suspension was diluted and used for a second infection as described above. After overnight incubation at 37°C, a single plaque was isolated and used as a stock for all subsequent manipulations.

The propagation procedure for bacteriophage 182 was modified from the agar
30 layer method of Swanstörn and Adams (1951). Briefly, the *Enterococcus* sp. PS strain was grown to stationary phase overnight at 37°C in TSB. The culture was then diluted 20 fold in TSB and incubated at 37°C until the $A_{540} = 0.2$. The suspension (15×10^7 Bacteria) was then mixed with 15×10^5 plaque forming units (pfu) to give a

ratio of 100-bacteria/pfu. After incubation of 15 min at 37°C, 7.5 ml of melted soft agar (TSB plus 0.6% agar) were added to the mixture and poured onto the surface of 150 mm TSA plates and incubated 16 hrs at 37°C. To collect the plate lysate, 20 ml of TSB were added to each plate and the soft agar layer was collected by scrapping off
5 with a clean microscope slide followed by vigorous shaking of the agar suspension for 5 min to break up the agar. The mixture was then centrifuged for 10 min at 4,000 rpm (2,830 xg) using a JA-10 rotor (Beckman) and the supernatant fluid (lysate) is collected and subjected to a treatment with 10 µg /ml of DNase I and RNase A for 30 min at 37°C. To precipitate the phage particles, the phage suspension was adjusted to
10 10% (w/v) of PEG 8000 and 0.5 M of NaCl followed by incubation at 4°C for 16 hrs. The phage was recovered by centrifugation at 4,000 rpm (3,500 xg) for 20 min at 4°C on a GS-6R table top centrifuge (Beckman). The pellet was resuspended with 2 ml of phage buffer (1 mM MgSO₄, 5 mM MgCl₂, 80 mM NaCl and 0.1% Gelatin). The phage suspension was extracted with 1 volume of chloroform and further purified by
15 centrifugation on a cesium chloride step gradient as described in Sambrook *et al.* (1989), using a TLS 55 rotor and centrifuged in an Optima TLX ultracentrifuge (Beckman) for 2 hrs at 28,000 rpm (67,000 xg) at 4°C. Banded phage was collected and ultracentrifuged again on an isopycnic cesium chloride gradient (1.45 g/ml) at 40,000 rpm (64,000 xg) for 24 hrs at 4°C using a TLV rotor (Beckman). The phages
20 were harvested and dialyzed for 4 hrs at room temperature against 4 L of dialysis buffer consisting of 10 mM NaCl, 50 mM Tris-HCl [pH 8] and 10 mM MgCl₂. Phage DNA was prepared from the phage suspension by adding 20 mM EDTA, 50 g/ml Proteinase K and 0.5% SDS and incubating for 1 hr at 65°C, followed by successive extractions with 1 volume of phenol, 1 volume of phenol-chloroform and 1 volume of
25 chloroform. The DNA was then dialyzed overnight at 4°C against 4 L of TE (10 mM Tris-HCl [pH 8.0], 1mM EDTA).

Example 12. DNA sequencing of the Bacteriophage 182 genome.

Four micrograms of phage DNA was diluted in 200 µl of TE (10 mM Tris,
30 [pH 8.0], 1 mM EDTA) in a 1.5 ml eppendorf tube and sonication was performed (550 Sonic Dismembrator, Fisher Scientific). Samples were sonicated under an amplitude of 3 µm with bursts of 5 s spaced by 15 s cooling in ice/water for 3 to 4

cycles. The sonicated DNA was then size fractionated by electrophoresis on 1% agarose gels utilizing TAE (1 x TAE is: 40 mM Tris-acetate, 1 mM EDTA [pH 8.0]) as the running buffer. Fractions ranging from 1 to 2 kbp were excised from the agarose gel and purified using a commercial DNA extraction system according to the instructions of the manufacturer (Qiagen), with a final elution of 50 µl of 1 mM Tris [pH 8.5].

The ends of the sonicated DNA fragments were repaired with a combination of T4 DNA polymerase and the Klenow fragment of *E. coli* DNA polymerase I, as follows. Reactions were performed in a reaction mixture (final volume, 100 µl) containing sonicated phage DNA, 10 mM Tris-HCl [pH 8.0], 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 50 µg/ml BSA, 100 µM of each dNTP and 15 units of T4 DNA polymerase (New England Biolabs) for 20 min at 12°C followed by addition of 12.5 units of the Klenow large fragment of DNA polymerase I (New England Biolabs) for 15 min at room temperature. The reaction was stopped by two phenol/chloroform extractions and the DNA was precipitated with ethanol and the final DNA pellet resuspended in 20 µl of H₂O.

Blunt-ended DNA fragments were cloned by ligation directly into the *Hinc* II site of the pKSII+ vector (New England Biolabs) dephosphorylated by treatment with calf intestinal alkaline phosphatase (New England Biolabs). A typical ligation reaction contained 100 ng of vector DNA, 2 to 5 µl of repaired sonicated phage DNA (50-100 ng) in a final volume of 20 µl containing 800 units of T4 DNA ligase (New England Biolabs) and was incubated overnight at 16°C. Transformation and selection of bacterial clones containing recombinant plasmids was performed in *E. coli* DH10β according to standard procedures (Sambrook *et al.*, 1989).

Recombinant clones were picked from agar plates into 96-well plates containing 100 µl LB and 100 µg/ml ampicillin and incubated at 37°C. The presence of phage DNA insert was confirmed by PCR amplification using T3 and T7 primers flanking the *Hinc* II cloning site of the pKS vector. PCR amplification of the potential foreign inserts was performed in a 15 µl reaction volume containing 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.02% gelatin, 1 µM primer, 187.5 µM each dNTP, and 0.75 units *Taq* polymerase (BRL). The thermocycling parameters were as follows: 2 min initial denaturation at 94°C for 2 min, followed by 20 cycles of 30 sec

denaturation at 94°C, 30 sec annealing at 58°C, and 2 min extension at 72°C, followed by a single extension step at 72°C for 10 min. Clones with insert sizes of 1 to 2 kbp were selected and plasmid DNA was prepared from the selected clones using the QIAprep™ spin miniprep kit (Qiagen).

5 The nucleotide sequence of the extremities of each recombinant clone was determined using an ABI 377-36 automated sequencer with two types of chemistry: ABI prism Big Dye™ primer cycle sequencing (21M13 primer: #403055)(M13REV primer: #403056) or ABI prism Big Dye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). To ensure co-linearity of the sequence data and
10 the genome, all regions of the phage genome were sequenced at least once from both directions on two separate clones. In areas that this criteria was not initially met, a sequencing primer was selected and phage DNA was used directly as sequencing template employing ABI prism BigDye™ terminator cycle sequencing ready reaction
15 kit.

Example 13. Bioinformatic management of primary nucleotide sequence.

Sequence contigs were assembled using Sequencher™ 3.1 software (GeneCodes). To close contig gaps, sequencing primers were selected near the edge of the contigs. Phage DNA was used directly as sequencing template employing ABI
20 prism BigDye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). The complete sequence of *Enterococcus* bacteriophage 182 is shown in Table 21.

A software program was used on the assembled sequence of bacteriophage 182 to identify all putative ORFs larger than 33 codons. The software scans the primary
25 nucleotide sequence starting at nucleotide #1 for an appropriate start codon. Three possible selections can be made for defining the nature of the start codon; I) selection of ATG, II) selection of ATG or GTG, and III) selection of either ATG, GTG, TTG, CTG, ATT, ATC, and ATA. This latter initiation codon set corresponds to the one reported by the NCBI([http://www.ncbi.nlm.nih.gov/htbin-](http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c)
30 [post/Taxonomy/wprintgc?mode=c](http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c)) for the bacterial genetic code. When an appropriate start codon is encountered, a counting mechanism is employed to count the number of codons (groups of three nucleotides) between this start codon and the

next stop codon downstream of it. If a threshold value of 33 is reached, or exceeded, then the sequence encompassed by these two codons is defined as an ORF. This procedure is repeated, each time starting at the next nucleotide following the previous stop codon found, in order to identify all the other putative ORFs. The scan is performed on all three reading frames of both DNA strands of the phage sequence. The predicted ORFs for bacteriophage 182 are listed in Tables 22 & 23. Sequence homology searches for each ORF were carried out using an implementation of BLAST programs. Downloaded public databases used for sequence analysis include:

- 10 (i) non-redundant GenBank (<ftp://ncbi.nlm.nih.gov/blast/db/nr.Z>),
- ii) Swissprot (<ftp://ncbi.nlm.nih.gov/blast/db/swissprot.Z>);
- iii) vector (<ftp://ncbi.nlm.nih.gov/blast/db/vector.Z>);
- iv) pdbaa databases (<ftp://ncbi.nlm.nih.gov/blast/db/pdbaa.Z>);
- v) staphylococcus aureus NCTC 8325 ([ftp://ftp.genome.ou.edu/pub/staph/staph-](ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa)
- 15 [1k.fa](ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa));
- vi) streptococcus pyrogenes
(ftp://ftp.tigr.org/pub/data/s_pneumoniae/gsp.contigs.112197.Z);
- vii) PRODOM
(ftp://ftp.toulouse.inra.fr/pub/prodom/current_release/prodom99.1.forblast.gz);
- 20 viii) DOMO (<ftp://ftp.infobiogen.fr/pub/db/domo/>);
- ix) TREMBL (ftp://www.expasy.ch/databases/sp_tr_nrdb/fasta/)

The results of the homology searches performed on the ORFs of bacteriophage 182 are shown in Tables 24 & 26.

25 **Example 14. Sub-Cloning of Bacteriophage 182 ORFs.**

Preparation of the shuttle expression vector

Expression preferably utilizes a shuttle expression vector which is arranged such that expression of the exogenous bacteriophage 182 ORF sequence is inducible. For example, the plasmid pND50 replicates in *E. coli*, *E. faecalis*, and *S. aureus*

30 (Yamagishi, J., Kojima, T., Oyamada, Y., Fujimoto, K., Hattori, H., Nakamura, S., and Inoue, M. 1996. *Antimicrob. Agents Chemother.* 40, 1157-1163). This plasmid can be modified by conventional techniques to insert the inducible arsenite promoter, derived from the shuttle vector pT0021, in which the firefly luciferase (*lucFF*)

expression is controlled by the *ars* promoter/operator from a *S. aureus* plasmid (Tauriainen, S., Karp, M., Chang, W and Virta, M. (1997). Recombinant luminescent bacteria for measuring bioavailable arsenite and antimonite. *Appl. Environ. Microbiol.* 63:4456-4461). This modified shuttle vector will contain the *ars* promoter, *arsR* gene and a cloning site for introduction of individual phage ORFs downstream from a shine-dalgarno sequence.

Other inducible regulatory sequences can be utilized instead of the arsenite-inducible system. An example is a nisin-inducible system. The *nisA* promoter activity is dependent on the proteins NisR and NisK, which constitute a two-component signal transduction system that responds to the extracellular inducer nisin. The nisin sensitivity and inducer concentration required for maximal induction varies among the strains, but is functional in *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Enterococcus faecalis*, and *Bacillus subtilis*. Significant induction of the *nisA* promoter (10- to 60-fold induction) can be obtained in all of the species. A vector containing this promoter was published as Eichenbaum Z, Federle MJ, Marra D, de Vos WM, Kuipers OP, Kleerebezem M, and Scott JR (1998) *Appl Environ Microbiol* 64, 2763-2769. Other vectors, e.g., plasmids, can also be utilized which will allow replication and transcription in *Enterococcus*.

Alternatively, a constitutive promoter can be used (e.g., the β -lactamase promoter is constitutive in *E. faecalis* – see ref. 1) to drive expression of the introduced ORF, and compare cell growth to control bacterial cells containing the parental vector lacking any introduced phage ORF. Recombinant plasmids are introduced into *E. faecalis* strain FA2-2 by electroporation, as previously described (Yamagishi, J., Kojima, T., Oyamada, Y., Fujimoto, K., Hattori, H., Nakamura, S., and Inoue, M. 1996. *Antimicrob. Agents Chemother.* 40, 1157-1163).

Cloning of ORFs with a Shine-Dalgarno sequence

ORFs with a Shine-Dalgarno sequence are selected for functional analysis of bacterial killing. Each ORF, from initiation codon to last codon (excluding the stop codon), will be amplified by PCR from phage genomic DNA. For PCR amplification of ORFs, each sense strand primer starts at the initiation codon and is preceded by a restriction site and each antisense strand starts at the last codon (excluding the stop codon) and is preceded by a different restriction site. The PCR product of each ORF will be gel purified and digested with the restriction enzymes with sites contained on

the PCR oligonucleotides. The digested PCR product is then gel purified using the Qiagen kit, ligated into the modified shuttle vector, and used to transform bacterial strain DH10 β . Recombinant clones are then picked and their insert sizes confirmed by PCR analysis using primers flanking the cloning site as well as restriction digestion.

- 5 The sequence fidelity of cloned ORFs will be verified by DNA sequencing using the same primers as used for PCR. In the cases that the verification of ORFs can not be achieved by one path of sequencing using primers flanking the cloning site internal primers will be selected and used for sequencing. Recombinant plasmids will be introduced into *E. faecalis* strain FA2-2 by electroporation, as previously described
- 10 (Yamagishi, J., Kojima, T., Oyamada, Y., Fujimoto, K., Hattori, H., Nakamura, S., and Inoue, M. 1996. *Antimicrob. Agents Chemother.* 40, 1157-1163).

Induction of gene expression from the *ars* promoter.

If an inducible promoter is used, e.g., the *ars* promoter, induction can be assessed, for example, in either of the two methods.

15 **1. Screening on agar plates**

The functional identification of killer ORFs can be performed by spreading an aliquot of *E. faecalis* transformed cells containing phage 182 ORF onto agar plates containing different concentrations of sodium arsenite (0; 2.5; 5; and 7.5 μ M). The plates are incubated overnight at 37°C, after which a growth inhibition of the ORF

- 20 transformants on plates that contain arsenite are compared to plates without arsenite.

2. Quantification of growth inhibition in liquid medium

- Cells containing different recombinant plasmids can be grown for overnight at 37°C in LB medium supplemented with the appropriate antibiotic selection. These are then diluted to the mid log phase ($OD_{540}=2$) with fresh media containing antibiotic
- 25 and transferred to 96-well microtitration plates (100 μ l/well). Inducer is then added at different final concentrations (ranging from 2.5 to 10 μ M) and the culture incubated for an additional 2 h at 37°C. The effect of expression of the phage 182 ORFs on bacterial cell growth is then monitored by measuring the OD_{540} and comparing the rate of growth to the culture not containing inducer. As positive controls for growth
- 30 inhibition, the *kilA* gene of phage lambda (Reisinger, GR., Rietsch, A., Lubitz, W. and Blasi, U. 1993 *Virology* #193: 1033-1036), and the *holin/lysin* genes of the *Staphylococcus aureus* phage Twort (Loessner, MJ., Gaeng, S., Wendlinger, G.,

Maier, SK. and Scherer, S. 1998. *FEMS Microbiology Letters* #162:265-274) were subcloned into the *ars* inducible vector. An aliquot of the induced and uninduced culture can also be plated out on agar plates containing an appropriate antibiotic selection but lacking inducer. Following incubation overnight at 37°C, the number of colonies is counted. Any ORF showing bacteriostatic activity will show a lower, but detectable, number of colonies on the agar plates when grown in the presence of inducer as compared to when grown in the absence of inducer. Any ORF showing bacteriocidal activity will show no colonies on the agar plates, when grown in the presence of inducer as compared to when grown in the absence of inducer.

10

REFERENCES

1. Cohen, M.L. (1992). *Science* **257**, 1050-1055.
2. Tenover, F.C. and McGowan Jr., J.E. (1998). Bacterial Infections of Humans. Epidemiology and Control. (A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 83-93.
3. Rusterholtz, K., and Pohlschroder, M. (1999). *Cell* **96**, 469-470.
4. Neu, H.C. (1992). *Science* **257**, 1064-1073.
5. Murray, B.E. (1990). *Clin. Microbiol. Rev.* **3**, 46-65.
6. Gray, B.M. (1998). Bacterial Infections of Humans. Epidemiology and Control. (A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 673 - 711.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning: A laboratory Manual*. Cold Spring Harbor Laboratory, New York. Cold Spring Harbor Laboratory Press.
7. Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J.
8. Rost B, I and Sander C. (1996). *Ann. Rev. Biophys. Biomol. Struct.* **25**, 113-136.
9. Garvey, K.J., Saedi, M.S., and Ito, J. (1985). *Gene* **40**, 311-316.
10. Pickett, G.G. and Peabody, D.S. (1993). *Nucl. Acids Res.* **21**, 4621-4626.
11. Gutiérrez, J., Vinos, J., Prieto, I., Mendez, E., Hermoso, J., and Salas, M. (1986). *Virology* **155**, 474-483.
12. Yoshikawa, H., Garvey, K.J., and Ito, J. (1985). *Gene* **37**, 125-130.
13. Martin, A.C., Lopez, R., Garcia, P. (1998). *J Bacteriol* **180**, 210-217.

14. Steiner, M., Lubitz, W., Blasi, U. (1993). J. Bacteriol. **175**, 1038-1042.
- Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., and Elledge, S.J. (1993). Genes Dev. **7**, 555-569.
- Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998). Mol Cell Biol. **18**,
5 2697-2711.
- Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura, Y., and Miki, Y. (1998). Genes, Chromosomes & Cancer **21**, 217-222.
- Endo, T.A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor
10 N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. (1997). Nature **387**, 921-924.
- Karimova, G., Pidoux, J., Ullmann, A., Ladant, D. (1998) Proc. Natl. Acad. Sci. **95**, 5752-5756.
- Sopta, M., Carthew, R.W., and Greenblatt, J. (1995) J. Biol. Chem. **260**, 10353-
15 10369.
- Qin, J., Fenyo, D., Zhao, Y., Hall, W.W., Chao, D.M., Wilson, C.J., Young, R.A. and Chait, B.T. (1997). Anal. Chem. **69**, 3995-4001.
- Swanström, M. and Adams, M.H. (1951). Proc. Soc. Exptl. Biol. Med. **78**, 372-375.

20

Example 15. Growth of *Streptococcus* bacteriophage Dp-1 and purification of genomic DNA.

- The *Streptococcus pneumoniae* R6 propagating strain (PS) (Tomasz, 1966) was used as host to propagate its respective phage Dp-1 (McDonnell et al.,
25 1975). (Alternatively, *Streptococcus (Diplococcus) pneumoniae* R36A could be used. Strain R36A is available from ATCC as #11733 or 27336. *Streptococcus pneumoniae* is also available from Felix d'Herelle Reference Center in Quebec, Canada as catalog number HER 1054. Other *S. pneumoniae* strains are also available from ATCC.)
- Two rounds of plaque purification of phage Dp-1 were performed on soft agar
30 essentially as described in Sambrook *et al.* (1989). Briefly, the *Streptococcus* R6 PS strain was grown overnight at 37°C in K-Cat media [K-Cat: 10 g Bacto casitone, 5 g Bacto tryptone, 1 g Yeast extract, 5g Potassium chloride, 0.2% Glucose, 30mM Potassium phosphate buffer [pH 8] and 250,000 Units Catalase per liter (Boehringer Mannheim #10683600). The culture was then diluted 20 fold in K-CAT and

incubated at 37°C until the $OD_{540} = 0.2$ (early log phase) with constant agitation. In order to obtain single plaques, Dp-1 phage was subjected to 10-fold serial dilutions using the phage buffer (100 mM Tris-HCl [pH 7.5], 100 mM NaCl and 10 mM $MgCl_2$) and 10 μ l of each dilution was used to infect 0.5 ml of the cell suspension.

- 5 After incubation of 15 min at 37°C, 2 ml of melted soft agar (K-CAT supplemented with 0.8% of agar) were added to the mixture and poured onto the surface of 100 mm K-CAT agar plates [K-CAT supplemented with 1.2 % of agar]. After solidification of the soft agar layer, an additional 5 ml of melted soft agar was added to visualize distinct plaques (Ronda et al., 1978). After overnight incubation at 37°C, a single
10 plaque was isolated, resuspended in 1 ml of phage buffer by end over end rotation for 2 hrs at room temperature, and the phage suspension was diluted and used for a second infection as described above. After overnight incubation at 37°C, a single plaque was isolated and used as a stock for all subsequent manipulations.

- The propagation procedure for bacteriophage Dp-1 was modified from the
15 agar layer method of Swanstörn and Adams (1951). Briefly, the R6 strain of *Streptococcus pneumoniae* was grown to stationary phase overnight at 37°C in K-CAT. The culture was then diluted 20 fold in K-CAT and incubated at 37°C until the $OD_{540} = 0.2$. The suspension (15×10^7 Bacteria) was then mixed with 15×10^5 plaque forming units (pfu) to give a ratio of 100-bacteria/pfu. After incubation of 15 min at
20 37°C, 7.5 ml of melted soft agar (K-CAT plus 0.8% agar) were added to the mixture and poured onto the surface of 150 mm K-CAT agar plates and incubated 16 hrs at 37°C. After solidification of the soft agar layer, 7.5 ml of melted soft agar were added to each plate. To collect the plate lysate, 20 ml of K-CAT media were added to each plate and the soft agar layers were collected by scrapping off with a clean microscope
25 slide followed by vigorous shaking of the agar suspension for 5 min to break up the agar. The mixture was then centrifuged for 10 min at 4,000 rpm (2,830 xg) using a JA-10 rotor (Beckman) and the supernatant (lysate) was collected and subjected to a treatment with 10 μ g /ml of DNase I and RNase A for 30 min at 37°C. To precipitate the phage particles, the phage suspension was adjusted to 10% (w/v) of PEG 8000 and
30 0.5 M of NaCl followed by incubation at 4°C for 16 hrs. The phage was recovered by centrifugation at 4,000 rpm (3,500 xg) for 20 min at 4°C on a GS-6R table top centrifuge (Beckman). The pellet was resuspended with 2 ml of phage buffer (100 mM Tris-HCl [pH 7.5], 100 mM NaCl and 10 mM $MgCl_2$). The phage suspension was extracted with 1 volume of chloroform and further purified by centrifugation on a
35 cesium chloride step gradient as described in Sambrook et al. (1989), using a TLS-55 rotor and centrifuged in an Optima TLX ultracentrifuge (Beckman) for 2 hrs at 28,000 rpm (67,000 xg) at 4°C. Banded phage was collected and ultracentrifuged again on an

isopycnic cesium chloride gradient (1.45 g/ml) at 40,000 rpm (64,000 xg) for 24 hrs at 4°C using a TLV rotor (Beckman). The phage was harvested and dialyzed for 4 hrs at room temperature against 4 L of dialysis buffer consisting of 10 mM NaCl, 50 mM Tris-HCl [pH 8] and 10 mM MgCl₂. Phage DNA was prepared from the phage suspension by adding 20 mM EDTA, 50 µg/ml Proteinase K and 0.5% SDS and incubating for 1 hr at 65°C, followed by successive extractions with 1 volume of phenol, 1 volume of phenol-chloroform and 1 volume of chloroform. The DNA was then dialyzed overnight at 4°C against 4 L of TE (10 mM Tris-HCl [pH 8.0], 1mM EDTA).

10

Example 16. DNA sequencing of the Bacteriophage Dp-1 genome.

Four micrograms of phage DNA was diluted in 200 µl of TE (10 mM Tris, [pH 8.0], 1 mM EDTA) in a 1.5 ml eppendorf tube and sonication was performed (550 Sonic Dismembrator, Fisher Scientific). Samples were sonicated under an amplitude of 3 µm with bursts of 5 sec spaced by 15 sec cooling in ice/water for 3 to 4 cycles. The sonicated DNA was then size fractionated by electrophoresis on 1% agarose gels utilizing TAE (1 x TAE is: 40 mM Tris-acetate, 1 mM EDTA [pH 8.0]) as the running buffer. Fractions ranging from 1 to 2 kbp were excised from the agarose gel and purified using a commercial DNA extraction system according to the instructions of the manufacturer (Qiagen), with a final elution of 50 µl of 1 mM Tris [pH 8.5].

The ends of the sonicated DNA fragments were repaired with a combination of T4 DNA polymerase and the Klenow fragment of *E. coli* DNA polymerase I, as follows. Reactions were performed in a reaction mixture (final volume, 100 µl) containing sonicated phage DNA, 10 mM Tris-HCl [pH 8.0], 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 50 µg/ml BSA, 100 µM of each dNTP and 15 units of T4 DNA polymerase (New England Biolabs) for 20 min at 12°C followed by addition of 12.5 units of the Klenow large fragment of DNA polymerase I (New England Biolabs) for 15 min at room temperature. The reaction was stopped by two phenol/chloroform extractions and the DNA was precipitated with ethanol and the final DNA pellet resuspended in 20 µl of H₂O.

Blunt-ended DNA fragments were cloned by ligation directly into the *Hinc* II site of the pKSII+ vector (New England Biolabs) dephosphorylated by treatment with calf intestinal alkaline phosphatase (New England Biolabs). A typical ligation reaction contained 100 ng of vector DNA, 2 to 5 µl of repaired sonicated phage DNA (50-100 ng) in a final volume of 20 µl containing 800 units of T4 DNA ligase (New England Biolabs) and was incubated overnight at 16°C. Transformation and selection

of bacterial clones containing recombinant plasmids was performed in *E. coli* DH10 β according to standard procedures (Sambrook *et al.*, 1989).

Recombinant clones were picked from agar plates into 96-well plates containing 100 μ l LB and 100 μ g/ml ampicillin and incubated at 37°C. The presence of phage DNA insert was confirmed by PCR amplification using T3 and T7 primers flanking the *Hinc* II cloning site of the pKS vector. PCR amplification of the potential foreign inserts was performed in a 15 μ l reaction volume containing 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.02% gelatin, 1 μ M primer, 187.5 μ M each dNTP, and 0.75 units *Taq* polymerase (BRL). The thermocycling parameters were as follows: 2 min initial denaturation at 94°C for 2 min, followed by 20 cycles of 30 sec denaturation at 94°C, 30 sec annealing at 58°C, and 2 min extension at 72°C, followed by a single extension step at 72°C for 10 min. Clones with insert sizes of 1 to 2 kbp were selected and plasmid DNA was prepared from the selected clones using the QIAprep™ spin miniprep kit (Qiagen).

The nucleotide sequence of the extremities of each recombinant clone was determined using an ABI 377-36 automated sequencer with two types of chemistry: ABI prism Big Dye™ primer cycle sequencing (21M13 primer: #403055)(M13REV primer: #403056) or ABI prism Big Dye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). To ensure co-linearity of the sequence data and the genome, all regions of the phage genome were sequenced at least once from both directions on two separate clones. In areas that this criteria was not initially met, a sequencing primer was selected and phage DNA was used directly as sequencing template employing ABI prism Big Dye™ terminator cycle sequencing ready reaction kit.

Example 17. Bioinformatic management of primary nucleotide sequence.

Sequence contigs were assembled using Sequencher™ 3.1 software (GeneCodes). To close contig gaps, sequencing primers were selected near the edge of the contigs. Phage DNA was used directly as sequencing template employing ABI prism BigDye™ terminator cycle sequencing ready reaction kit (Applied Biosystems; #4303152). The complete sequence of *Streptococcus* bacteriophage Dp-1 is shown in Table 28.

A software program was used on the assembled sequence of bacteriophage Dp-1 to identify all putative ORFs larger than 33 codons. The software scans the primary nucleotide sequence starting at nucleotide #1 for an appropriate start codon. Three possible selections can be made for defining the nature of the start codon; I) selection of ATG, II) selection of ATG or GTG, and III) selection of either ATG,

GTG, TTG, CTG, ATT, ATC, and ATA. This latter initiation codon set corresponds to the one reported by the NCBI(<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c>) for the bacterial genetic code. When an appropriate start codon is encountered, a counting mechanism is employed to count the number of codons (groups of three nucleotides) between this start codon and the next stop codon downstream of it. If a threshold value of 33 is reached, or exceeded, then the sequence encompassed by these two codons is defined as an ORF. This procedure is repeated, each time starting at the next nucleotide following the previous stop codon found, in order to identify all the other putative ORFs. The scan is performed on all three reading frames of both DNA strands of the phage sequence. The predicted ORFs for bacteriophage Dp-1 are listed in Tables 29 and 30, and Fig. 6.

Sequence homology searches for each ORF were carried out using an implementation of BLAST programs. Downloaded public databases used for sequence analysis include:

- (i) non-redundant GenBank (<ftp://ncbi.nlm.nih.gov/blast/db/nr.Z>),
- ii) Swissprot (<ftp://ncbi.nlm.nih.gov/blast/db/swissprot.Z>);
- iii) vector (<ftp://ncbi.nlm.nih.gov/blast/db/vector.Z>);
- iv) pdbaa databases (<ftp://ncbi.nlm.nih.gov/blast/db/pdbaa.Z>);
- v) staphylococcus aureus NCTC 8325 (<ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa>);
- vi) streptococcus pyogenes (ftp://ftp.tigr.org/pub/data/s_pneumoniae/gsp.contigs.112197.Z);
- vii) PRODOM (ftp://ftp.toulouse.inra.fr/pub/prodom/current_release/prodom99.1.forblast.gz);
- viii) DOMO (<ftp://ftp.infobiogen.fr/pub/db/domo/>);
- ix) TREMBL (ftp://www.expasy.ch/databases/sp_tr_nrd/fasta/)

The results of the homology searches performed on the ORFs of bacteriophage Dp-1 are shown in Table 31.

Example 18. Sub-Cloning of Bacteriophage Dp-1 ORFs.

Preparation of the shuttle expression vector

Expression preferably utilizes a shuttle expression vector which is arranged such that expression of the exogenous bacteriophage Dp-1 ORF sequence is inducible. For example, the plasmid pLSE4 replicates in *E. coli*, and *S. pneumoniae* (Diaz and Garcia, 1990). This plasmid can be modified by conventional techniques to insert the inducible arsenite promoter, derived from the shuttle vector pT0021, in which the

firefly luciferase (*lucFF*) expression is controlled by the *ars* promoter/operator from a *S. aureus* plasmid (Tauriainen, S., Karp, M., Chang, W and Virta, M. (1997).

Recombinant luminescent bacteria for measuring bioavailable arsenite and antimonite. *Appl. Environ. Microbiol.* 63:4456-4461). This modified shuttle vector will contain the *ars* promoter, *arsR* gene and a cloning site for introduction of individual phage ORFs downstream from a shine-dalgarno sequence.

Other inducible regulatory sequences can be utilized instead of the arsenite-inducible system. An example is a nisin-inducible system. The *nisA* promoter activity is dependent on the proteins NisR and NisK, which constitute a two-component signal transduction system that responds to the extracellular inducer nisin. The nisin sensitivity and inducer concentration required for maximal induction varies among the strains, but is functional in *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Enterococcus faecalis*, and *Bacillus subtilis*. Significant induction of the *nisA* promoter (10- to 60-fold induction) can be obtained in all of the species. A vector containing this promoter was published as Eichenbaum Z, Federle MJ, Marra D, de Vos WM, Kuipers OP, Kleerebezem M, and Scott JR (1998) *Appl Environ Microbiol* 64, 2763-2769. Other vectors, e.g., plasmids, can also be utilized which will allow replication and transcription in *Streptococcus*.

Alternatively, a constitutive promoter can be used to drive expression of the introduced ORF, and compare cell growth to control bacterial cells containing the parental vector lacking any introduced phage ORF. Recombinant plasmids are introduced into *S. pneumoniae* R6 as previously described (Diaz and Garcia, 1990)

Cloning of ORFs with a Shine-Dalgarno sequence

ORFs with a Shine-Dalgarno sequence are selected for functional analysis of bacterial killing. Each ORF, from initiation codon to last codon (excluding the stop codon), will be amplified by PCR from phage genomic DNA. For PCR amplification of ORFs, each sense strand primer starts at the initiation codon and is preceded by a restriction site and each antisense strand starts at the last codon (excluding the stop codon) and is preceded by a different restriction site. The PCR product of each ORF will be gel purified and digested with the restriction enzymes with sites contained on the PCR oligonucleotides. The digested PCR product is then gel purified using the Qiagen kit, ligated into the modified shuttle vector, and used to transform bacterial strain DH10 β . Recombinant clones are then picked and their insert sizes confirmed by PCR analysis using primers flanking the cloning site as well as restriction digestion. The sequence fidelity of cloned ORFs will be verified by DNA sequencing using the same primers as used for PCR. In the cases that the verification of ORFs can not be achieved by one path of sequencing using primers flanking the cloning site

internal primers will be selected and used for sequencing. Recombinant plasmids will be introduced into *S. pneumoniae* R6 as previously described (Diaz and Garcia, 1990).

Induction of gene expression from the *ars* promoter.

- 5 If an inducible promoter is used, e.g., the *ars* promoter, induction can be assessed, for example, in either of the two methods.

1. Screening on agar plates

- The functional identification of killer ORFs can be performed by spreading an aliquot of *S. pneumoniae* transformed cells containing phage Dp-1 ORFs onto agar plates containing different concentrations of sodium arsenite (0; 2.5; 5; and 7.5 μ M).
10 The plates are incubated overnight at 37°C, after which a growth inhibition of the ORF transformants on plates that contain arsenite are compared to plates without arsenite.

2. Quantification of growth inhibition in liquid medium

- Cells containing different recombinant plasmids can be grown for overnight at
15 37°C in LB medium supplemented with the appropriate antibiotic selection. These are then diluted to the mid log phase ($OD_{540}=0.2$) with fresh media containing antibiotic and transferred to 96-well microtitration plates (100 μ l/well). Inducer is then added at different final concentrations (ranging from 2.5 to 10 μ M) and the culture incubated for an additional 2 hrs at 37°C. The effect of expression of the phage Dp-1 ORFs on
20 bacterial cell growth is then monitored by measuring the OD_{540} and comparing the rate of growth to the culture not containing inducer. [As positive controls for growth inhibition, the *kilA* gene of phage lambda (Reisinger, GR., Rietsch, A., Lubitz, W. and Blasi, U. 1993 *Virology* #193: 1033-1036), and the *holin/lysin* genes of the
Staphylococcus aureus phage Twort (Loessner, MJ., Gaeng, S., Wendlinger, G.,
25 Maier, SK. and Scherer, S. 1998. *FEMS Microbiology Letters* #162:265-274) can be subcloned into the *ars* inducible vector. An aliquot of the induced and uninduced culture can also be plated out on agar plates containing an appropriate antibiotic selection but lacking inducer. Following incubation overnight at 37°C, the number of colonies is counted. Any ORF showing bacteriostatic activity will show a lower, but
30 detectable, number of colonies on the agar plates when grown in the presence of inducer as compared to when grown in the absence of inducer. Any ORF showing full bacteriocidal activity will show no colonies on the agar plates, when grown in the presence of inducer as compared to when grown in the absence of inducer.

35 **REFERENCES**

15. Cohen, M.L. (1992) *Science* **257**, 1050-1055.

16. Tenover, F.C. and McGowan Jr., J.E. (1998) Bacterial Infections of Humans. Epidemiology and Control.(A.S. Evans and P.S. Brachman, eds.) Plenum Medical Book Company, New York, N.Y. pp. 83-93.
- 5 17. Rusterholtz, K., and Pohlschroder, M. (1999) *Cell* 96, 469-470.
18. Klugman, K.P. (1990) *Clin. Microbiol. Rev.* 3, 171-196.
19. Fenoll, A., Martin Bourgon, C., Munoz, R., Vicioso, D., Casal, J. (1991) *Rev. Infect. Disease* 13, 56-60.
- 10 20. Jorgensen, J.H., Doern, G. V., Maher, L. A., Howell, A. W., Redding, J. S. (1990) *Antimicrob. Agents Chemother.* 34, 2075-2080.
- 15 21. Neu, H.C. (1992) *Science* 257, 1064-1073.
- Hsueh, P. R., Wu, J. J., Hsiue, T. R. (1996) *J Formos Med Assoc* 5, 364-371.
 - Garcia, P., Martin, A.C., and Lopez, R. (1997) *Microbial Drug Res.* 3, 165-176.
 - 20 • Martin, A.C., Lopez, R., and Garcia, P. (1996) *J. Virol.* 70, 3678-3687.
 - Sheehan, M.M., Garcia, J.L., Lopez, R., and Garcia, P. (1997) *Mol. Microbiol.* 25, 717-725.
 - 25 • Kodaira, M., Biswas, S.B., and Kornberg, A. (1983) *Mol. Gen. Genet.* 192, 80-96.
 - Maki, S. and Kornberg, A. (1988) *J. Biol. Chem.* 263, 6547-6554.
 - 30 • Tsuchihashi Z, Kornberg A. (1990) *Proc. Natl. Acad. Sci. USA.* 87, 2516-2520.
 - Lee, S.H. and Walker, J.R. (1987) *Proc Natl Acad Sci USA* 84, 2713-2717.
 - Smidt, C.R., Steinberg, F.M., Rucker, R. (1991) *Proc Soc Exp Biol Med* 197, 19-26.
 - 35 • Frank, D.W, (1997) *Mol Microbiol.* 26, 621-629.
 - Nardese, V., Gutlich, M., Brambilla, A., Carbone, M.L.(1996) *Biochem Biophys Res Commun* 218, 273-279.
 - 40 • Mancini, R., Saracino, F., Buscemi, G., Fischer, M., Schramek, N., Bracher, A., Bacher, A., Gutlich, M., Carbone, M.L. (1999) *Biochem Biophys Res Commun* 255, 521-527.
 - 45 • Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: A laboratory Manual.* Cold Spring Harbor Laboratory, New York. Cold Spring Harbor Laboratory Press.

22. Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J.
23. Rost B.I and Sander C. (1996) *Ann. Rev. Biophys. Biomol. Struct.* **25**, 113-136.
- 5 24. Garvey, K.J., Saedi, M.S., and Ito, J. (1985) *Gene* **40**, 311-316.
25. Pickett, G.G. and Peabody, D.S. (1993) *Nucl. Acids Res.* **21**, 4621-4626.
- 10 26. Gutiérrez, J., Vinos, J., Prieto, I., Mendez, E., Hermoso, J., and Salas, M. (1986) *Virology* **155**, 474-483.
27. Yoshikawa, H., Garvey, K.J., and Ito, J. (1985) *Gene* **37**, 125-130.
- 15 28. Martin, A.C., Lopez, R., Garcia, P. (1998) *J Bacteriol* **180**, 210-217.
29. Steiner, M., Lubitz, W., Blasi, U. (1993) *J. Bacteriol.* **175**, 1038-1042.
- 20 • Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., and Elledge, S.J. (1993). *Genes Dev.* **7**, 555-569.
- Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998) *Mol Cell Biol.* **18**, 2697-2711.
- 25 • Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura, Y., and Miki, Y. (1998) *Genes, Chromosomes & Cancer* **21**, 217-222.
- Endo, T.A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. (1997) *Nature* **387**, 921-924.
- 30 • Karimova, G., Pidoux, J., Ullmann, A., Ladant, D. (1998) *Proc. Natl. Acad. Sci.* **95**, 5752-5756.
- 35 • Sopta, M., Carthew, R.W., and Greenblatt, J. (1995) *J. Biol. Chem.* **260**, 10353-10369.
- Qin, J., Fenyo, D., Zhao, Y., Hall, W.W., Chao, D.M., Wilson, C.J., Young, R.A. and Chait, B.T. (1997) *Anal. Chem.* **69**, 3995-4001.
- 40 • Tomasz, A. (1966) *Journal of Bacteriology* **91**, 1050-1061.
- McDonnell, M., Ronda, LC and Tomasz, A. (1975) *Virology* **63**, 577-582.
- 45 • Ronda C., Lopez, R., Tomasz, A. and Portoles A. (1978) **26**, 221-225.

- Swanström, M. and Adams, M.H. (1951) Proc. Soc. Exptl. Biol. Med. 78, 372-375.
 - Díaz E and García JL. (1990) Gene 90, 163-167.
- 5 • Tauriainen, S., Karp, M., Chang, W and Virta, M. (1997). Recombinant luminescent bacteria for measuring bioavailable arsenite and antimonite. Appl. Environ. Microbiol. 63:4456-4461.

10 All patents and publications mentioned in the specification are indicative of the levels of skill of those skilled in the art to which the invention pertains. All references cited in this disclosure are incorporated by reference to the same extent as if each reference had been incorporated by reference in its entirety individually.

15 One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The specific methods and compositions described herein as presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

20 It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. For example, those skilled in the art will recognize that the invention may suitably be practiced using a variety of different bacteria, bacteriophage, and sequencing methods within the general descriptions provided.

25 The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is not intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

In addition, where features or aspects of the invention are described in terms of Markush groups or other grouping of alternatives, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group or other group. For example, 5 if there are alternatives A, B, and C, all of the following possibilities are included: A separately, B separately, C separately, A and B, A and C, B and C, and A and B and C. Thus, for example, for the bacteria and phage specified herein, the embodiments expressly include any subset or subgroup of those bacteria and/or phage. While each such subset or subgroup could be listed separately, for the sake of brevity, such a 10 listing is replaced by the present description.

Thus, additional embodiments are within the scope of the invention and within the following claims.

Table 1

Phages against human and animal pathogenic bacteria

5

I. Pathogen name	Phage name	II. Catalog#	Origin/reference
<i>Acinetobacter calcoaceticus</i>	A3/2 A10/45 A36 B9GP B ₉ PP BS46 E13 E14 531		Felix d'Herelle Reference Centre, Quebec, Quebec
	Ap3 P78		J. Bacteriol 1984. 157: 179-183 J. Gen. Microbiol 1986.132: 2633-2636
<i>Acinetobacter haemolyticus</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Acinetobacter johnsonii</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Acinetobacter sp.</i>	BP1		J.Virol.1968.2:716-722
	G4, HP2, HP3 & HP4		Can.J.Microbiol.1966.12:1023-1030 & J.Virol.1974.13:46-52 & Arch.Virol.1994.135:345-354
	A1, A4, A9 & 196		Arch.Virol.1994.135:345-354
	HP1		Can.J.Microbiol.1966.12:1023-1030
	A19, A23, A29, A31, A33, A34, A3759 & 2845		J.Microsc (Paris) 1973.16:215-224 & CR.Hebdo Seances Acad.Sci.Ser D.Sci Natur(Paris)278:1907-1909 & Arch.Virol.1994.135:345-354 & Rev.Can.Biol.1970.29:317-320
<i>Actinobacillus actinomycetecomitans</i>			FEMS Microbiol Lett 1994. 119:329-337.

			Infect. Immun. 1982. 35: 343-349
			Mol.Gen.Genet 1998.258: 323-325
	Aaφ247		Oral Micriol. Immunol 1997.12: 40-46
<i>Actinomyces viscosus</i>		43146-B1	The American Type Culture Collection
			Infect.Immun.1985.48:228-233
			Infect.Immun.1988.56:54-59
			Plasmid 1997.37:141-153
<i>Aeromonas hydrophila</i>	PM2** & PM3		FEMS Microbiol.Lett. 1990.57:277-282
	Ach1 Ach2 PM4 PM5 PM6 T7-ah		Felix d'Herelle Reference Centre, Quebec, Quebec

<i>Aeromonas salmonicida</i>	3 25 29 31 32 40RR _{2st} 43 51 56 59.1 65 Asp37		Felix d'Herelle Reference Centre, Quebec, Quebec
	55R.1		Can. J. Microbiol. 1983. 29: 1458-1461
<i>Alteromonas espejiana</i>	PM2**	27025-B1	The American Type Culture Collection
<i>Asticcacaulis biprosthecum</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Asticcacaulis excentricus</i>		15261-B1 15261-B2 15261-B3	The American Type Culture Collection
	φAc21 φAc24		
<i>Azotobacter vinelandii</i>		12518-B1 12518-B4 12518-B5 12518-B9 12518-B10 13705-B1	The American Type Culture Collection
	A14 A21 A31 A41 PAV1		
<i>Azotobacter sp.</i>			Virology 1972.49:439-452
<i>Bacteroides fragilis</i>	Bf-1		Rev. Infect. Dis. 1979. 1: 325-336
	B40-8		FEMS Microbiol. Lett. 1991. 66: 61-67
	HSP40		Appl. Environ. Microbiol. 1989. 55: 2696-2701
	phiA1		Zentralbl.bakteriol.1972.222:57-63
<i>Bdellovibrio bacteriovorus</i>	MAC-1		J. Gen. Microbiol. 1987. 133: 3065-3070
<i>Bdellovibrio sp.</i>	VL-1		J.Virol.1973.12:1522-1533
<i>Bordetella bronchiseptica</i>	214		Zh.Mikrobiol.Epidemiol.Immuno. 1987.5:9-13

<i>Bordetella parapertussis</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
			Mol. Gen. Mikrobiol. Virusol. 1988.4: 22-25
			Zh.Mikrobiol.Epidemiol.Immuno. 1987.5:9-13
	41405		Zh.Mikrobiol.Epidemiol.Immuno. 1987.5:9-13
<i>Brucella abortus</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
	10/I 24/II 212/XV	23448-B1 23448-B2 23448-B3 17385-B1 17385-B2	The American Type Culture Collection
	BK-2, TB & Fj**		Zh.Mikrobiol.Epidemiol.Immunobiol.1983.2: 48-52
	R/c & R/O		Dev. Biol. Stand. 1984.56: 55-62
	R/c		Dev. Biol. Stand. 1984.56: 55-62
<i>Brucella canis</i>	BK-2	23456-B1	The American Type Culture Collection
<i>Brucella melitensis</i>	Wb		Zentralbl.Veterinarmed.1975.22:866-867

	Fi** & TB		Zh.Mikrobiol.Epidemiol.Immunobiol.1983.2: 48-52
<i>Brucella sp.</i>			Can. J. Vet. Res. 1989.53: 319-325
			Res. Vet. Sci. 1988. 44: 45-49
	R		Zh.Mikrobiol.Epidemiol.Immunobiol.1983.2: 48
<i>Campylobacter coli</i>		43133-B1	The American Type Culture Collection
		43134-B1	
<i>Campylobacter coli</i> (Cont'd)	18	43135-B1	The American Type Culture Collection
	19	43136-B1	
	20		
<i>Campylobacter jejuni</i>	1	35918-B1	The American Type Culture Collection
	2	35919-B1	
	3	35920-B1	
	4	35921-B1	
	5	35918-B2	
	6	35920-B2	
	7	35922-B2	
	8	35923-B1	
	9	35924-B1	
	10	35925-B1	
	11	35925-B2	
	12	35922-B2	
	13	35924-B2	
	14	35922-B3	
	17	43133-B1	
	18	43134-B1	
	19	43135-B1	
	20	43136-B1	
<i>Campylobacter</i> (<i>Helicobacter</i>) <i>pylori</i>	HP1		J. Med. Microbiol.1993. 38: 245-249
<i>Chlamydia psittaci</i>	Chp1**		J. Gen. Virol. 1989. 70: 3381-3390
<i>Clostridium</i> <i>acetobutylicum</i>	CAK-1		J.Bacteriol.1993.175:3838-3843

<i>Clostridium botulinum</i>			Nucleic Acids Res.1990.18:1291
			Bioch.Biophys.res.Commun.1990.171.1304-1311
			Microbiol.immunol.1981.25:915-927
			J.Vet.Med.Sci.1992.54:675-684
	CE β & CE γ		
<i>Clostridium difficile</i>	41 & 56		J. Clini.Microbiol. 1985.21:251-254

<i>Clostridium perfringens</i>			Rev.Can.Biol.1977.36:205-215
			FEMS Microbiol.Lett. 1990.54:323-326
<i>Clostridium sporogenes</i>	59 70 71 72S 72L	8074-B1 17886-B1 17886-B3 17886-B4 17886-B5 17886-B6	The American Type Culture Collection
<i>Clostridium tetani</i>	A & B		Rev.Can.Biol.1978.37:43-46
<i>Corynebacterium diphtheriae</i>			Vopr.Virusol.1986.31:577-584
<i>Corynebacterium pseudotuberculosis</i>	NN	12319-B1	The American Type Culture Collection
<i>Corynebacterium sp</i>	DLC 2921/49	12052-B1	The American Type Culture Collection

<i>Enterococcus faecalis</i>	42	19948-B1	The American Type Culture Collection
<i>Enterococcus faecium</i>	124 133	19950-B1 19953-b2 19953-B1	The American Type Culture Collection

<i>Escherichia coli</i>		11303-B14	The American Type Culture Collection
		11303-B10	
		11303-B21	
		8677-B1	
		11303-B13	
		13706-B4	
<i>Escherichia coli</i> (Cont'd)		15766-B1	The American Type Culture Collection
		15766-B1	
		1242-B5	
		15669-B2	
		15767-B1	
		11303-B16	
		27-65-B1	
		25065-B2	
	C204	15669-B1	
	E1	15597-B1	
	π^{**}	21816-B1	
	π^{**}	23724-B9	
	FCZ	15593-B1	
	fd**	25404-B1	
		29746-B1	
		23631-B1	
		25868-B1	
		25298-B1	
		25298-B2	
		11303-B37	
		11303-B24	
		11303-B26	
	Ifi**	11303-B27	
		11303-B28	
		11303-B29	
		11303-B30	
		11303-B33	
		11303-B31	
		11303-B25	
		11303-B35	
		11303-B34	
	MS2**	11303-B36	
	MU9	11303-B32	
	Mu-1	13706-B5	
	Ox6	11303-B1	
	P1**	11303-B2	
	P4 sid, **	11303-B3	
	Q- β **	11303-B4	
	R17**	35060-B1	
	Z1K/1	35060-B2	
	ZJ/2	35060-B3	
		11303-B5	
		11303-B6	
		11303-B7	
		11303-B38	
		12141-B1	

<i>Escherichia coli</i> (Cont'd)		11303-B20	The American Type Culture Collection
		11303-B17	
		11303-B15	
		11303-B11	
	547	11303-B18	
	UV1	13706-B2	
	UV47	23724-B2	
	UV375	23724-B1	
	$\alpha 3^{**}$	23724-B3	
	λ . **	23724-B4	
	λ C-17	23724-B5	
	λ sus P-3	23724-B6	
	λ sus R-5	23724-B7	
	λ sus J-6	23724-B8	
	λ sus O-8	35860-B1	
	λ sus A-11	13706-B3	
	λ ind'	15597-B2	
	$\phi 92$	13706-B1	
	ϕR	49696-B1	
	$\phi V-1$		
	$\phi X174^{**}$		
	$\phi Xcs70am-3$		
	G4** & ϕK^{**}		Biochim.Biophysica Acta.1992.1130:277-288
	BF23**		J.Bacteriol.1977.129:265-275
	Mu1		J.Ultrastruct.Res.1966.14:441-448
	Hpl7		J.Mol.Biol.1991.218:705-721
	K3** & Ox2**		FEBS Lett.1987.215:145-150
	Rb18**, Rb51 & Rb69**		J.Bacteriol.1990.172:180-186
	H1**, H3, H8, K9, K18 & Ox1		Mol.Gen.Genet.1990.221:491-494
	M1**, Tula** & Tulb**		J.Mol.Biol.1987.196:165-174
	K10		J.Bacteriol.1979.140:680-686
	Qsr'		J.Bacteriol.1985.162:256-262
	B278		J.Gen.Microbiol.1988.134:1333-1338
	phi 80**		FEMS Microbiol.Lett.1994.119:71-76
	phi m173		Genetika 1985.21:673-675
	tf-1		J.Gen.Microbiol.1987.133:953-960
	P4 & phiR73		Mol.Microbiol.1995.18:201-208
	I ₂ -2		J.Gen.Microbiol.1982.128:2797-2804
	PRD1		Virology 1990.177:445-451
	K3hx		Mol.Gen.Genet.1987.206:110-115
	933J** & 933W**		Infect.Immunity.1986.53:135-140
	H19-B**		J.Bacteriol.1987.169:4308-4312
	Tcp-111		Zentralbl.Bakteriol.Mikrobiol.Hyg.1988.270:41-51

<i>Escherichia coli</i> (Cont'd)	N4**	Vet.Microbiol.1992.30:203-212
	Phi 80 trp	Ann.Inst.Pasteur.1971.120:121-125
	Obeta 1	J.Bacteriol.1978.133:172-177
	P1CM	J.Gen.Microbiol.1978.107:73-83
	PA-2**	J.Bacteriol.1990.172:1660-1662
	186**	Mol.Gen.Genet.1982.187:87-95
	186.IX.B	Mol.Microbiol.1992.6:2629-2642
	21**	Virology 1983.129:484-489
	P4**	MicrobiolRev.1993.57:683-702
	82**	J.Biol.Chem.1987.262:11721-11725
	PSP3	J.Bacteriol.1996.178:5668-5675
	HK022**	Nucleic Acids Res.1994.22:354-356
	D108**	Nucleic Acids Res.1986.14:3813-3825
	Rb49	J.Mol.Biol.1997.267:237-249
	Ike**	J.Mol.Biol.1985.181:27-39
	P22dis	Mol.Gen.Genet.1978.166:233-243
	N15**	J.Bacteriol.1996.178:1484-1486
	If1**	Proc.R.Soc.Lond.B.Biol.Sci.1991.245:23-30
	Stx2Phi-I & Stx2Phi-II	Infect.Immun.1998.66:4100-4107
	18	Virology 1987.156:122-126
	X	J.Gen.Microbiol.1981.126:389-396
	AC3	Mol.Microbiol.1991.5:715-725

	BW-1 C-1 E920g Esc-7-11 H19J Haiti HK243 Ia K20 K30 KL, M Mu** O103 O157:H7 PID pt1 PilHa PR64FS PR772 SS4 β 4Q λ vir** Ω 8 09-1 92		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Haemophilus influenzae</i>	HP1**		Nucleic Acids Res. 1996.24:2360-2368
	S2**		Gene 1997. 196: 139-144
<i>Halobacterium cutirubrum</i>	S45		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Halobacterium halobium</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
			Can.J.Microbiol.1982.28:916-921
<i>Halobacterium salinarium</i>			Biol.Chem.Hoppe Seyler 1994.375:747-757

<i>Klebsiella oxytoca</i>	tf-1		J.Gen.Microbiol.1987.133:953-960
<i>Klebsiella pneumoniae</i>	60	23356-B1	The American Type Culture Collection
	92	23357-B1	
	K19Q		
	FC3-1 & FC3-9		
	FC3-10		FEMS Microbiol.Lett.1991.67:291-297
<i>Klebsiella sp.</i>	K11**		Mol.Gen.Genet. 1990.221:283-286
<i>Leptospira sp.</i>	LE1, LE3 & LE4		Res.Microbiol.1990.141:1131-1138
<i>Listeria monocytogenes</i>	243	23074-B1	The American Type Culture Collection
	197,1313 & 9425		Appl.Environ.Microbiol.1997.63:3374-3377
	H387 & H387-A		Appl.Environ.Microbiol.1993.59:2914-2917
	5775,6223 &12682		APMIS.1993.101:160-167
	2389, 2671, 4211 & 2685		Intervirology 1994.37:31-35 & Zentralbl.Bakteriol.Mikrobiol.Hyg.1986.261:1 2-28
	4b, 4ab, 4g & 3c		Ann.Microbiol. (Paris) 1977.128:185-198
	A118, A500 & A511**		Mol.Microbiol. 1995.16:1231-1241-992
	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 16, 17, 19 & 20		Ann.Microbiol. (Paris) 1979.130B:179-189
	1/2a, 1/2b, 3c, 4ab, 6a & 6b		Clin.Invest.Med.1984.7:229-232
	φLMUP35 2685		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Listeria innocua</i>	4211		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Micrococcus luteus</i>		4698-B1	The American Type Culture Collection
		4698-B4	
	N3	4698-2	
	N4	4698-B3	
	N8		
<i>Micrococcus luteus</i>	N17		Can.J.Microbiol. 1979.25:1027-1035
<i>Mycobacterium smegmatis</i>	BK-3	27203-B1	The American Type Culture Collection
	Bo1**	27204-B1	
	Bo 6	27205-B1	
	Bo 6II	27205-B2	
	Bo 6III	27205-B3	
	Mc-2	607-B6	
	Mc-4	607-B7	
	NN	11727-B1	
	Phagus lacticola	11759-B1	
	R1	607-B1	

Legendre Leo Roy Sedge	HER 317 HER 330 HER 333 HER 335 HER 334 HER 331 HER 316	Felix d'Herelle Reference Centre, Quebec, Quebec
		Mol.Microbiol.1993.7:395-405
		J.Mol.Biol.1998.279:143-164
		Proc.Natl.Acad.Sci USA.1988.84:2833-2837
		Mol.Biol.Rep. 1981.30:11-15
		Proc.Natl.Acad.Sci.USA 1997.94:10961-10966
	29M, 31M, 122, 154, 37, 29D, 46, 139,110, 141, 74D, AG1 & DS6A	Arch.Virol.1993.133:39-49 & Am.Rev.Respir.Dis.1975.112:17-22
<i>Mycobacterium fortuitum</i>	Bo 4 Bo 7	23052-B1 27207-B1 27207-B2 The American Type Culture Collection

<i>Mycobacterium leprae</i>			Ann.Microbiol. (Paris) 1982.133:93-97
<i>Mycobacterium tuberculosis</i>	DS6A	25618-B1 25618-B2 4243-B1	The American Type Culture Collection
	110, 139 & 33D		Arch.Virol.1993.133:39-49
	AG1,GS4E, BG1, PH & BK1		The Biology of Mycobacteria.Academic Press,Toronto 1982 (Ratledge & Stanford) 1982.309-351
<i>Mycobacterium sp</i>	Phagus pellegrini NN B1	11760-B1 11761-B1 23239-B1	The American Type Collection Culture

	TM4, ph60, ph72, PhAE39, phAE40 & Bxb1		Microbiology 1995.141:1173-1181
	C2		Experientia 1969.25:1112-1113
	18 & 115		J.Gen.Virol.1987.68:949-956
	63		Gruzlica 1968.36:617-622
	phlei & butyricum		J.Gen.Virol.1975.29:235-238
	MyF3P-59a		Z.Allg.Mikrobiol.1968.8:29-37
	Bo2a		J.Gen.Virol.1973.20:75-87
	D4,D28 & D32		J.Exptl.Med.1966.123:327-340
	HC		J.Bacteriol.1963.86:608-609
<i>Mycobacterium vaccae</i>	B5	15483-B1	The American Type Culture Collection
<i>Mycobacterium phlei</i>	NN Bo 2 Bo 2h Bo 3	11728-B1 11758-B1 27086-B2 27086-B1	The American Type Culture Collection
<i>Mycoplasma arthritidis</i>	MAV1**		Infect.Immunity.1995.63:4016-4023
<i>Mycoplasma hyorhinis</i>	Hr-1		Arch.Virol.1983.77:81-85
<i>Mycoplasma pneumoniae</i>	Br-1		Arch.Virol.1983.75:1-15
<i>Mycoplasma pulmonis</i>			Plasmid 1995. 33: 41-49
<i>Mycoplasma sp.</i>			J.Gen.Microbiol.1985:131:3117-3126
			J. Virol.1986.59:584-590
			Gene 1994. 141: 1-8

		Microbios 1990. 64: 111-125
		Infection& Immunity 1995. 63: 4016-4023
		Med.Biol.1982.60:116-120
MV-L2 &		Arch.Virol.1979.61:289-296
		Acta.Virol.1978.22:443-450
		J.Gen.Virol.1979.42:315-322
		Virology 1973.55:118-126

			Science 1971.173:725-727
<i>Neisseria perflava</i>			J.Clin.Microbiol.1976. 4:87-91
<i>Nocardia erythropolis</i>	φC		J.Gen.Virol.1974.23:247-254
	φEC		J.Bacteriol.1976.126:1104-1107
<i>Pasteurella multocida</i>	B225		Arch.Exp.Veterinarmed.1981.35:433-436
	B939a		Am.J.Vet.Res.1978.39:1565-1566
	Nos.115, 32, 967 & 1075		Vet.Med.Nauki. 1977.14:33-36
<i>Propionibacterium acnes</i>	NN	29399-B1	The American Type Collection Culture

<i>Pseudomonas aeruginosa</i>	2	12175-B1	The American Type Culture Collection
	2A	12175-B2	
	2B	12175-B3	
	11	12175-B4	
	16	14205-B1	
	24	14206-B1	
	27	14207-B1	
	44	14208-B1	
	73	14209-B1	
	95	14210-B1	
	109	14211-B1	
	113	14212-B1	
	249	14213-B1	
	B3	14214-B1	
	Hoff 2	15692-B1	
	Hoff 3	14203-B1	
	Pa	14204-B1	
	Pb	12055-B1	
	PB-1	12055-B2	
	Pc	15692-B3	
	Pf	12055-B3	
	PP7**	25102-B1	
		15692-B2	
7 & 31			Felix d'Herelle Reference Centre, Quebec, Quebec
	Pf3**		J.Virol.1983.47:221-223
	φ-MC		Can.J.Microbiol.1969.15:1179-1186
	Pf1**		J.Mol.Biol.1991.218:349-364
	PR4**		J.Gen.Virol.1979.43:583-592
	A7		J.Bacteriol.1992.174:2407-2411
	KF1		J.Biochem.1983.93:61-71
	CTX**		Mol.Microbiol.1993.4:1703-1709
	φ2**		J.Virol.1977.24:135-141

	<p>φKZ, 21, φNZ, PMN17, PTB80, 68, PB-1, E79, 16, 109, 352, 1214, F8, 71, 337, M4, φC17, SL2, B17, Li-24, φmnP78, PS17**, φ1, 73, M6, Li-2, 7, φmnF82, PTB2, PTB20, PTB42, φKF77, 31, PTB21, 119x, φPLS27, B3, 258, Hw12, PM57, PM62, PM105, 148, PM681, 198, 218, 222, 242, 246, PC131, φC11, SL5, D3112**, Jb19, F7, PM69, PM13, PM61, PM113, φ240, 249 & 269</p>		ddd
--	---	--	-----

<i>Pseudomonas aeruginosa</i> (Cont'd)	297, 309, 318, 11,		Arch.Virol.1993.131:141-151
---	-----------------------	--	-----------------------------

<i>Pseudomonas cepacia</i>			Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Pseudomonas fragi</i>		27362-B1 27363 B1	The American Type Culture Collection
	wy		
<i>Pseudomonas phaseolicola</i>	φ6		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Pseudomonas putida</i>	gh-1	12633-B1	The American Type Culture Collection
<i>Pseudomonas syringae</i>		40492-B1 21781-B1	The American Type Culture Collection
	φ-6		
<i>Pseudomonas sp.</i>	PPs-G3	49780-B1	The American Type Culture Collection
<i>Salmonella bareilly</i>	Sab 2		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Salmonella enteritidis</i>	1, 2, 3 & 6		Epidemiol. Infect. 1995.114:227-236
	2a, 3a, 4a, 5a, 6a, 7a, 8a, 9a, 15, 19, 20 & 21**		Vet. Med. Nauki. 1975.12:55-60
<i>Salmonella newington</i>	Epsilon 34		J. Struct. Biol. 1995.115:283-289
<i>Salmonella newport</i>		27869-B1 27869-B2	The American Type Culture Collection
	16-19		
			Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Salmonella paratyphi</i>		19940-B1 12176-B1	The American Type Culture Collection
	Paratyphoid A		
	Jersey		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Salmonella senftenberg</i>	SasL1, SaL2, Sal 3, SaL4, SaL5 & SasL6		Indian J. Med. Res. 1997.105:47-52
<i>Salmonella typhimurium</i>	P22** SL-1	19585-B1 40282	The American Type Culture Collection
	MB78**		J. Virol. 1982.41: 1038-1043
	SE1		J. Gen. Microbiol. 1986.132:1035-1041
	LT2		Virology 1971.45:835-636
	ES18**		Virology 1970.42:621-632
	L**		J. Virol. 1985.56:1034-1036

	P1CM clr-100		Mol.Gen.Genet.1975.138:113-126
	F22		Genet.Res.1986.48:139-143
	Fels 1		J.Gen.Virol.1978.38:263-272
	Fels 2		Genet.Res.1986.48:139-143
	Px		Mol.Gen.Genet.1970.108:184-202
	P1kc		Virology 1974.60:503-514
	A3 & A4		J.Bacteriol. 1987.169:1003-1009
	HT		Genet.Res.1976.27:315-322
<i>Salmonella typhimurium</i> (Cont'd)	IRA		J.Basic Microbiol. 1990.30:707-716
	Mud1		Mol.Gen.Genet. 1986.202:327-330
	P22 (cir4-1, cir5-1 & cir6-1)		Mol.Gen.Genet. 1984.198:105-109
	BF23**		Mol.Gen.Genet.1976.147:195-202
	Kb1		J.Bacteriol.1974.117:907-908
	P221dis		J.Gen.Virol.1978.41:367-376
	PRD1**		Virology 1990.177:445-451
	I ₂ -2**		J.Gen.Microbiol.1982.128:2797-2804
	tf-1		J.Gen.Microbiol.1987.133:953-960
<i>Salmonella typhosa/typhi</i>	X**		J.Gen.Microbiol.1981.126:389-396
	8	19937-B1	The American Type Culture Collection
	23	19938-B1	
	25	19939-B1	
	46	19942-B1	
	53	19943-B1	
	163	19946-B1	
	175	19947-B1	
	VII	27870-B1	
	ViVI	27870-B2	
	O1		Felix d'Herelle Refrence Centre, Quebec, Quebec
<i>Salmonella sp.</i>	VIII		Chung Hua Liu Hsing Ping H.T.C.1992.13:288
	j2		J.Gen.Microbiol.1983.129:3395-33400
	P3	25957-B1	The American Type Culture Collection
	P4**	25957-B2	
	P9a	25957-B3	
	P9c	25957-B4	
	P10	25957-B5	
	102	19945-B1	
	Chi (x)	9842-B1	
	R34	97541	
	MG40		Virology 1968.34:521-530
	P14		Microb.Pathog.1990.8:393-402
	PSP3		Virology 1992.188:414
	Ike**		Zentralbl.Bakteriol.1976.234:294-304
<i>Sphaerotilus natans</i>	P27 & 9NA		J.Virol.1986.12:921-931
	SN1		Appl.Environ.Microbiol.1979.37:1025-1030

<i>Shigella dysenteriae</i>	P2 ø80	23351-B1 11456b 11456a-B1	The American Type Culture Collection
<i>Shigella flexneri</i>	D20	12661-B1	The American Type Culture Collection
	SfII**		Mol.Microbiol.1997.26:939-950
	SfV**		Gene 1997.22:217-227
	Sf6**		Mol.Microbiol.1995.18:201-208
	SfX		Gene 1993.129:99-101
<i>Shigella sonnei</i>	C16**		
	Ufa		Mol..Biol (Mosk) 1977.11:323-331
<i>Shigella sp</i>	37	23354-B1	The American Type Culture Collection
<i>Spiroplasma citri</i>	SpV1		Plasmid 1993.29:193-205
<i>Spiroplasma sp.</i>	SpV1-R8A2B		Nucleic Acids Res. 1990.18:1293
	SpV3		Isr.J.Med.Sci.1987.23:429-433
	Sp V4		J.Bacteriol.1987.169:4950-4961
<i>Staphylococcus albus</i>			Staphylococci & Staphylococcal Infections.1997. Voll:503-508 (Karger,Basel)

<i>Staphylococcus aureus</i>		27702-B1	The American Type Culture Collection
		27703-B1	
		27704-B1	
		23360-B1	
		23361-B1	
	15	27705-B1	
	17	27712-B1	
	29	27690-B1	
	42D**	27691-B1	
	42E	27692-B1	
	47	27693-B1	
	52	27694-B1	
	52A	27695-B1	
	53	27696-B1	
	54	27697-B1	
	55	27698-B1	
	71	27699-B1	
	75	27693-B2	
	77	27700-B1	
	79	27701-B1	
	80	27706-B1	
	81	27707-B1	
	83A	27708-B1	
	84	33742	
	85**	33741-B1	
	88	15565	
	92	19685-B1	
	5504'	11987-B1	
	K	11988-B1	
	P1	15752-B1	
	P14		
	UC18		

	HER 101 HER 239 HER 283 HER 49	Felix d'Herelle Reference Centre, Quebec, Quebec
Twort**		
φ11**		J.Bacteriol.1988.170:2409-2411
φ13** & φ42**		J.Gen..Microbiol.1989.135:1679-1697
L54a**		J.Bacteriol.1986.166:385-391
80α**		Can.J.Microbiol.1996.43:612-616
94,95 & 96		J.Clin.Microbiol.1988.26:2395-2401
φ131, A ₃ & A ₅		Staphylococci & Staphylococcal Infections.1997. Vol1:503-508 (Karger,Basel)
Phi PVL**		Gene 1998.215:57-67
<i>Staphylococcus carnosus</i>	BaSTC2	Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Staphylococcus epidermidis</i>	1a, 2b, 3a, 4b, 5a, 6b, 7b, 8c, 9a, 10a, 11b, 12a & 13b	Can.J.Microbiol.1988.34:1358-1361
	41, 63, 118II, 138, 245, 336, 392 & 550	Res.Virol.1994.145:111-121
<i>Staphylococcus saprophyticus</i>	1154A, 1405, 1314, 1139 & 1259	Res.Virol.1990.141: 625-635 & Res.Virol.1994.145:111-121
<i>Staphylococcus sp.</i>	Phi 812, Phi 131, SK311 & U16	Virology 1998.246:241-252
<i>Streptococcus faecalis</i>	VD13	HER44 Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Streptococcus faecium</i>	PE1	Zentralbl.Bakteriol.1975.231:421-425
<i>Streptococcus oralis</i>	Cp-1** & Cp- 7**	FEMS Microbiol.Lett.1989.65:187-192

<i>Streptococcus pneumoniae</i>	Cp-1**	HER223	Felix d'Herelle Reference Centre, Quebec, Quebec
	Cp-1**, Cp-5**, Cp-7**, Cp-9**, ω -1 & ω -2		J.Virol.1981.40:551-559 & Eur.J.Biochem.1979.101:59-64 & Microbial Drug Resistance 1997.3:165-176
	HB-623 & HB-746		J.Virol.1990.64:5149-5155
	EJ-1**		J.Bacteriol.1992.174:5516-5525
	Dp-2 & Dp-4		J.Virol.1978.26:221-225
	Dp-1		Virology 1975.63:577-582
	ω -3 & ω -8		J.Virol.1976.19:659-667
	304		J.Bacteriol.1980.141:1298-1304
	HB-1, HB-2, HB-3**, HB-4, HB-5 & HB-6		J.Bacteriol.1979.138:618-624
<i>Streptococcus pyogenes</i>	T12**		Mol. Microbiology. 1997#23:719-728
	A-1	12202-B1	The American Type Culture Collection
	A-6	12203-B1	
	A-25	12204-B1	
	Kjem	14918	
<i>Streptococcus sp./Enterococcus</i>	1	HER 339	Felix d'Herelle Reference Centre, Quebec, Quebec
	182	HER 80	
	VD1884	HER 323	
	1A	12169-B1	The American Type Culture Collection
	1B	12170-B1	
	NN	21597-B1	
	42	19948-B1	
	118	19951-B2	
	120	19952-B1	
<i>Veillonella rodentium</i>	N2		Antonie Van Leeuwenhoek 1989.56:263-271
<i>Vibrio cholerae</i>	Psi 92		Intervirology 1993.36:237-244
	VCB-1,2,3 & 4		J.Infection 1998.36:131
	CP-T1**		J.Virol.1984.51:163-169
	VSK		FEMS Microbiol.Lett.1996.145:17-22
	Phi138		J.Virol.1986.57:960-967
	Phi149		J.Virol.1985.140:217-223
	Fs-2**		Microbiology 1998.144:1901-1906

	e4 e5 X29 β κ 13 14 16 24 32 57		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Vibrio cholerae</i> (Cont'd)	138 145 149 163 N-4 S-5 S-20 M-4 D-10 I II III IV V	14100-B1 14100-B2 14100-B30 14100-B4 51352-B1 51352-B2 51352-B3 51352-B4 51352-B5 51352-b6 51352-B7 51352-B8 51352-B9 51352-B10	The American Type Culture Collection
<i>Vibrio costicola</i>	UTAK		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Vibrio eltor</i>	e ₄		J.Gen.Virol.1987.68:1411-1416
<i>Vibrio natrigens</i>	nt1, nt6		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Vibrio</i> <i>parahaemolyticus</i>	KVP40** VF33 VP1 ϕ 60 ϕ HAWI-5 ϕ PEL8C-1		Felix d'Herelle Reference Centre, Quebec, Quebec
<i>Vibrio</i> sp.	α 3a		Felix d'Herelle Reference Centre, Quebec, Quebec
	NN ph1	11985-B1 51582-B1	The American Type Culture Collection
	Phil49		J.Virol.1987.61:3999-4006
<i>Veillonella rodentium</i>	N2		Antonie V.Leeuwenhoek.1989.56:263-271

<i>Yersinia enterocolitica</i>	1 2 3 4 5 6 7 8 9 φYeO3-12		Felix d'Herelle Reference Centre, Quebec, Quebec
	I, IV & VIII		Zentralbl. Bakteri. Mikrobiol. Hyg. 1982. 253:1 02
<i>Yersinia pestis</i>	R S Y	23208-B1 11593-B1 23053-B1	The American Type Culture Collection
	II		Zh. Mikrobiol. Epidemiol. Immunobiol. 1990. 11 :9
<i>Yersinia pseudotuberculosis</i>	PST**	23207-B1	The American Type Culture Collection
<i>Yersinia sp.</i>	RD2		Mol. Gen. Mikrobiol. Virusol. 1990. 8:18-21

xxx)

Table 2

>Bacteriophage 77, complete genome sequence, 41708 nucleotides

```

1      gatcaaaaata cttgggggaac ggtaggggag taaacttcgc gataatttta aaaatttcag
61     tataaccccc ctcttataac cattttaagg caggtagatg aatggagatt atagtcgatg
121    aaaatttagt gcttaaagaa aaagaaaggc tacaagtatt atataaagac atacctagca
181    ataaatttaa agtagttgat ggtttaatta ttcaagcagc aaggctacgt gtaatgcttg
241    attacatgtg ggaagacata aaagaaaaag gtgattatga tttatttact caatctgaaa
301    aggcgcacc atattgaaagg gaaagaccag tagccaaact atttaatgct agagatgctg
361    catatcaaaa aataatcaaa caattatcgg atttattgcc cgaagagaaa gaagacacag
421    aaacgccatc tgatgattac ctatgattag taataaatac gttgatgaat atataaattt
481    gtggaaacaa ggaaagataa ttttaaataa agaaagaatt gatctcttta attatctaca
541    aaaacatata tattcacgag atgatgtata ttttgatgaa cagaaaaatcg aggattgtat
601    caaatttatt gaaaaatggg attttccaac attaccattt caaaggttta tcatagctaa
661    tatatttctt atagataaaa atacagatga agctttcttt acagaatttg ctattttcat
721    gggacgtgga ggcgggaaaa acggtctaat aagtgtctatt agtgatttct tttctacgcc
781    cttacacgga gttaaagaat atcacatctc cattgttgct aatagtgaag atcaagcaaa
841    aacatcgttt gatgaaatca gaaccgtttt aatggataac aaacgaaaaa agacgggtaa
901    aacgcctaaa gctccttatg aagttagtaa agcaaaaata ataaaccgtg caactaaatc
961    ggtttattcga tataacacat caaacacaaa aaccaaagac ggtggacgtg aggggtgtgt
1021   tatttttgat gaaattcatt atttctttgg tcctgaaatg gtaaacgtca aacgtggtgg
1081   attaggtaaa aagaaaaata gaagaacgct ttatataagt actgatggtt ttgttagaga
1141   gggttatatc gatgcaatga agcacaatat tgcaagtgtt ttaagtggca aggttaaaaa
1201   tagtagattt ttgtcttttt attgtaagtt agacgatcca aaagaagttg atgacagaca
1261   gacgtgggaa aaggcgaaac caatgttaca taaccgttta tcagaatacg ctaaaacact
1321   gctaagcacg attgaagaag aatataacga tttaccattc aaccgttcaa ataagcccg
1381   attcatgact aagcgaatga atttgcttga agttgacctt gaaaaagtaa tagcaccatg
1441   gaaagaataa ctagcgacta atagagagat accaaattta gataatcaaa tgtgtattgg
1501   tgggttagac ttgcaaaaaa ttcgagattt tgcaagtgtt gggctattat tccgaaaaaa
1561   cgatgattac atttggttag gacattcgtt tgtaagacaa gggtttttgg atgatgtcaa
1621   attagaacct cctattaaag aatgggaaaa aatgggatta ttgaccattg tcatgatga
1681   tttcatatga attgaatata tagttgattg gtttttaaag gctagagaaa aatatgggct
1741   tgaaaaagtc atagctgata attatagaac tgatattgta agacgtgctg ttgaggatgc
1801   tggcataaaa cttgaagtac ttgaaatcc aaaagcaata catggattac ttgcaccacg
1861   tatcgataca atgtttgcga aacataacgt aatatatgga gacaatcctt tgatgcgttg
1921   gtttactaat aatgttgctg taaaaatcaa gccgtagtga aataaagagt atacaaaaa
1981   agatgaagtc agacgtaaaa cggatggatt catggctttt gttcacgcat tatatagagc
2041   agacgatata gtagacaaag acatgtctaa agcgtctgat gcattaatga gtatagattt
2101   ctaatagagg aggtgagaca tgagttattc agaaaagata tttaaaacta ggaagatat
2161   taacgaggtt cttgatttag atatgataga agatctatca caacaagcgt atgtgaaacg
2221   tttagcgatt gatagttgta ttgaatttgt tgcgcgagct gtcgctcaa gtcattttta
2281   agtattggaa ggtaatagaa ttcaaaaaga tgatgtttac tacaagttaa atataaaacc
2341   aaatactgac ttatcaagcg atagtttttg gcaacaagtt atataataac taatttatga
2401   taacgaggtt ttaatcgtag taagtgcagc caaagaatta cttatcgagc atagctttta
2461   cagagaagag tacgctttgt atgatgatat attcaaagat gtaacggtta aagattatac
2521   ttatcaacgt actttcacaa tgcaagaggt catatattta aagtacaaca acaataaagt
2581   gacacacttt gtagaaagtc tattcgaaga ttacgggaaa atattcggaa gaatgatagg
2641   tgcacaatta aaaaactatc aaataagagg gattttgaaa tctgcctcta gcgcatatga
2701   cgaaaagaat atagaaaaat tacaagcgct cacaataaaa ttattcaata cttttaataa
2761   aatcaacta gcaatcgcgc ctttgataga aggttttgat tatgaggaaat tatctaattg
2821   tggtaagaat agtaacatgc ctttttctga attgagtgag ctaatgagag atgcaataaa
2881   aatgtttgct ttgatgattg gtatacctcc aggtttgatt tacggagaaa cagctgattt
2941   ggaaaaaac acgcttgat ttgagaagtt ctgtttaaca cctttattaa aaaagattca
3001   gaacgaatta aacgcgaaac tcataacaca aagcatgtat ttgaaagata caagaataga
3061   aattgtcggg gtgaataaaa aagaccact tcaatatgct gaagcaattg acaacttgt
3121   aagttctggt tcatttacaa ggaatgaggt gcggattatg ttagggtgaag aaccatcaga
3181   caatcctgaa ttagacgaat acctgattac taaaaactac gaaaaagcta acagtgggtg
3241   aatgatgaa aaagaaaaag atgaaaacac tttgaaaggt ggtgatgaag atgaaagcgg
3301   agattaaagg cgtcatcgtt tccaacgaag ataaatgggt ttacgaaatg cttggtatgg
3361   attcgacttg tcttaagat gttttaacac aactagaatt tagtgatgaa gatgttgata
3421   ttataattaa ctcaaatggt ggtaacctag tagctggtag tgaatatat acacatttaa
3481   gagctcataa aggcacaagt aatgttcgta tcacagcaat agcagcaagt gcggcatcgc
3541   ttatcgcaat ggctggtgac cacatcgaaa tgagtcgggt tgctagaatg atgattcaca

```

3601 atccttcaag tattgcgcaa ggagaagtga aagatctaaa tcatgctgca gaaacattag
3661 aacatggttg tcaataaatg gctgagggat atgctggttag agctggtaaa aacaaacaag
3721 aacttataga aatgatggct aaggaaacgt ggctaaatgc tgaatgaagcc attgaacaag
3781 gttttgcgga tagtaaaatg tttgaaaacg acaatatgca aattgtagca agcgatacac
3841 aagtgttatc gaaagatgta ttaaatcgtg taacagcttt ggtaagtaaa acgccagagg
3901 ttaacattga tattgacgca atagcaataa aagtaattga aaaaataaat atgaaagaaa
3961 aggaatcaga aatcgatggt gcagatagta aattatcagc aaatggattt tcaagattcc
4021 ttttttaata caaaaatagg aggtcataaa atgactataa atttatcgga aacattcgca
4081 aatgcgaaaa acgaatttat taatgcagta aacaacgggt aaccgcaaga aagacaaaat
4141 gaattgtacg gtgacatgat taaccaacta tttgaagaaa cttaattaca agcaaaaagca
4201 gaagctgaaa gaggtttctag tttacctaaa tcagcacaaa ctttgagtgc aaaccaaaaga
4261 aatttcttta tggatatcaa taagagtgtt ggatataaag aagaaaaact tttaccagaa
4321 gaaacaattg atagaatctt cgaagattta acaacgaatc atccattatt agctgactta
4381 ggtattaaaa atgctggttt gcgtttgaag ttcttaaaat ccgaaacttc tggcgtgggt
4441 gtttggggtt aaatctatgg tgaattataa ggtcaattag atgctgctgt cagtgaagaa
4501 acagcaattc aaaaataatt gacagcgttt gttgttttac caaaagattt aatgatattt
4561 ggtcttgcgt ggattgaaag atttgttcgt gttcaaatcg aagaagcatt tgcagtggcg
4621 cttgaaactg cgttcttaaa aggtactggt aaagaccacac cgattggctt aaaccgtcaa
4681 gtacaaaaag gtgtatcggt aactgatggt gcttatccag agaaaagaaga acaaggtagc
4741 cttacatttg ctaatcgcgt cgtacggtt aatgaattga cgaaagtgtt taaataccac
4801 tcaactaacg agaaaggtaa atcagtagcg gttaaaggta atgtaacaat ggttggtaat
4861 ccgtccgagt cttttgaggt tcaagcacag tatacacatt taaatgcaaa tggcgtatat
4921 gttactgctt tactatttaa tttgaatggt attgagtcta cagttcaaga agcaggtaag
4981 gtttttaacgt acgttaagg tctatatgat ggttattag ctgggtggat taatgttcag
5041 aaatttaaaag aaacacttgc gttagatgat atggatttat acactgcaaa acaatttgct
5101 tacggcaaaag cgaagataa taaagtgtg gctgtttgga aattagattt aaaaggacat
5161 aaaccagctt tagaagatac cgaagaaaca ctataaaatt ttatgaggtg ataaaaaggt
5221 gaaatttaaa gttgttagag aatttaaga catagagcac aatcaacaca agtacaaggt
5281 aggggagttg tatccagctg aagggtataa caatcctcgt gttgaattgt tgacaaatca
5341 aatcaaaaat aagtagcaga aagtttatat cgtaccttta gataagctga caaaacaaga
5401 attattagaa ctatgcgaat cattacaaaa aaaagcgtct agttcaatgg ttaaaagtga
5461 aatcatcgac ttattgaatg gtgaagacaa tgacgattga tgatttgctt gtcaaattta
5521 aatcacttga aaagattgac cataattcag aggatgagta cttaaaagcag ttgttaaaaa
5581 tctcgtacga gcgtataaaa aatcagtcgc gaggttttga attagagaat ttaaataggc
5641 aagaattgat acttatacgc gctagatag cttatcaaga ttatttagaa cacttcaacg
5701 acaattacag acctgaataa atagattttt cgttatctct aatggaggta tcagaagatg
5761 aagaaagtgt ttaagaaacc tagaattaca actaaacgtt taaatacgcg tgttcatatt
5821 tataaggtata ctgaaaaataa tgggtccgaa gctggagaaa aagaagaaaa attattatat
5881 agctgttggg cgagtattga tgggtgtctg ttacgtgaat tagaacaagc tatctcaaac
5941 ggaacgcgaa atgacattaa attgtatat cgtgatccgc aagggtgatta ttaccaggt
6001 gaagaacatt atcttgaatc tgaatcaaga tatttcaaaa atcgtttgaa tataaagcaa
6061 gtatcaccag atttggataa taaagacttt attatgattc gcggaggata tagttcatga
6121 gtgtgaaagt gacaggtgat aaagcattag aaagagaatt agaaaaacat tttggcataa
6181 aagagatggt aaaagttcaa gataaggcgt taatagctgg tgctaaggta attgttgaag
6241 aaataaaaaa acaactcaaa ccttcagaag actcaggagc actgattagt gagattggct
6301 gtactgaacc tgaatggata aaggggaaac gtactgttac aattaggttg cgtgggcttt
6361 ttgaacgatt tagaatagta catttaattg aaaatggcca tggttgagaa aagtcaggaa
6421 aatttgcata acctaagctt atgggtggga ttaatagagc aataagacaa gggcaaaata
6481 agtattttga gacgctaaaa agggagttga aaaaattgtg attgatattt tgtacaaagt
6541 tcatgaagtg attagtcaag acagaattat tagagagcac gtaaatatca ataattatga
6601 gttcaataaa taccctaagt taaaagatac tgatgtacct tttattgtta ttgacgatat
6661 cgacgaccca atacctacaa cttatactga cggagatgag tgtgcatata gttatatgtt
6721 ccaaatagat gttttgttta agtacaatga tgaatataat gcgagaatca caagaaataa
6781 gatattcaat cgcattcaaa agttattatg gtctgaacta aaaaagggaa atgtttcaaa
6841 tggaaaaacc gaatatatag aagaatttaa aacatataga agctctcgcg ttacagagg
6901 cattttttat aaggaggaaa attaaatggc agtaaaacat gcaagtgcgc caaaggcgta
6961 tattaacatt actggtttag gtttcgctaa attaacgaaa gaaggcgcgc aattaaaata
7021 tagtgatatt acaaaaaaa gaggattaca aaaaattggt gttgaaactg gtggagaact
7081 aaaaacagct tatgctgatg gcggtccaat tgaatcaggg aatacagagc gagaaggtaa
7141 aatctcatta caaatgcagc cgttccctaa agagattcgc aaaattgttt ttaatgaaga
7201 ttatgatgaa gatggcggtt acgaagagaa acaaggtaaa caaaacaatt acgtagctgt
7261 atgggttcaga caagagcgta aagacgggtac atttagaaca gttttattac cttaaagttat
7321 gttttacaat cctaaaatcg atggagaac ggctgagaaa gattgggatt tctcaagtga
7381 agagggtgaa ggtgaggcac ttttcccttt agttgataat aaaaagtcag tacgtaagta
7441 tatctttgat tcagctaaac tgacaaatca tgatggagac ggtgaaaaag gcgaagaggc
7501 tttcttaaaag aaaatttttg gcgaagaata tactggaac gtgacagagg gtaacgaaga
7561 aactttgtaa caaaaccggc ttcatcgga actgcggtta atgctggtta tataccagat
7621 agcattaaaa cacttaaagt tggcgacaca tacgatttaa atgttgtagt agagccatct

7681 aatcaaagta agttattgaa atacacaaca gatcaaacga atattgtatc aatcaatagt
 7741 gatgggtcaag ttacttgcga agcacaaggc attgtctacgg ttaaaagcaac agttggtaat
 7801 atgagtgaca ctataacaat aaatgtagaa gcataagagg gggcaacccc tctattttat
 7861 ttgaaaataa ggagagtatt ataaaatggc aaaattaaaa cgtaacatta tcaattagt
 7921 agaagatcca aaagcaaatg aaactaaatt acaaacgtac ttaacaccac acttcatttc
 7981 atttgaattt gtatacgaag caatggattt aatcgatgat attgaggagc aaaaatgcac
 8041 gatgaagcca agagaaatcg ctgacagatt gatggatag gttgtaaaaa ttacgataa
 8101 ccaattcaca gttaaagacc taaaagaacg tatgcatgca cctgatggaa tgaatgcact
 8161 tcgtgaacaa gtgattttca ttactcaagg tcaacaaact gaggaacta gaaattttat
 8221 ccagaacatg aaataaagcc tgaagattta acatataaag caatgttgaa aaatatggat
 8281 actctcatga tggacttaat tgaaaatggc aaagacgcta acgaagtttt aaaaatgcca
 8341 tttcattatg tgctttccat atatcaaaat aaaaataatg acatttctga agaaaaagca
 8401 gaggtcttaa ttgatgcatt ttaaccttaa ccgtttgggt aggggttatt tttgaaactt
 8461 ttttagaaag gaggtaaaa atggggagaa gaataaaagg tttatctata ggtttggact
 8521 tagatgcagc aaatttaaat agatcatttg cagaaatcaa acgaaacttt aaaaacttaa
 8581 attctgactt aaaatttaaca ggcaacaact tcaaatatac cgaaaaatca actgtagatt
 8641 acaaaacaaag gattaaagaa cttgatggaa ctatcacagg ttataagaaa aacgttgatg
 8701 atttagccaa gcaatatgac aaggtatctc aagaacaggc cgaaaaacgt gcagaagctc
 8761 aaaagttacg acaagaatat aacaaacaag caaatgagct gaattattta gaaagagaat
 8821 tacaaaaaac ttcatataca ggttctgacg tcaaaaaagc tcaagttgaa gctcaagaaa
 8881 tggcagaaag tggctgggga aaaaccagta aagtttttga aagtatggga cctaaattaa
 8941 caaaaaatgg tgatggttta aaatccattg gtaaaagggt gatgatgggt gtaactgcac
 9001 ctgttttagt tattgcagca gcatcaggaa aagcttttgc agaagttgat aaaggtttag
 9061 atactgtttac tcaagcaaca ggcgcaacag gcagtgaatt aaaaaaattg cagaactcat
 9121 ttaagatgtt ttatggcaat ttccagcag atgctgaaac tgttgggtgga gttttaggag
 9181 aagttataac aaggttaggt ttacacagta aagaacttga aaatgccaca ggtcatttct
 9241 tgaatttcag tcataataca ggttctgacg gtgtgcaagc cgtacagtta attaccgtg
 9301 caatgggcga tgcaggtatc gaagcaagtg aatatcaaa tgttttggat atggtagcaa
 9361 aagcggcgca agctagtggg ataagtgttg atacattagc tgatagtatt actaaatcag
 9421 gcgctccaat gagagctatg ggctttgaga tgaagaatc aattgcttta ttctctcaat
 9481 gggaaaagtc aggcgttaat actgaaatag cattcagtggt ttgaaaaaaa gctatatcaa
 9541 attggggtaa agctgggttaa aaccacaagag aagaatttaa gaagacatta gcagaattg
 9601 aaaaagacgc ggtatagct agcgcacaa gtttagcgat tgaagcattt ggtgcaagg
 9661 caggctcctga tttagcagac gctattaaag gtggctgctt tagttatcaa gaatttttaa
 9721 aaactattga agattcccaa ggcacagtaa accaaacatt taaagattct gaaagtggct
 9781 ccgaaagatt taaagtagca atgaataaat taaaattagt aggtgtgat gtagggctt
 9841 ctattgaaag tgcgtttgct cccgtaattg aagaattaat caaaaagcta tctatagcgg
 9901 ttgattgggt ttccaattta agtgatgggt ctaaaagatc aattgttatt ttcagtggta
 9961 ttgctgctgc aattggctct gtagtttttg ggtaggtgac atttataagt acaattggca
 10021 atgcagtaac tgtattagct ccatgtttag ctagtattgc aaaggctggg ggattgatta
 10081 gttttttatc gactaaaagta cctatattag gaactgtctt cacagcttta actgggtcaa
 10141 ttggcattgt attaggtgta ttggctgggt tagcagtcgc atttacaatt gcttataaga
 10201 aatctgaaac atttagaatt ttgtttaatg gtgcaattga aagtgttaaa caaacattta
 10261 gtaattttat tcaattttat caacctttcg ttgattctgt taaaaacatc tttaaacaag
 10321 cgtatcagc aatagttgat ttgcacaaag atatttggag tcaaatcaat ggattcttta
 10381 atgaaaacgg aatttccatt gttcaagcac ttcaaaatat atgcaacttt attaaagcga
 10441 tatttgaatt tatttttaatt ttgttaatta aaccaattat gttcgcgat tggcaagtga
 10501 tgcaatttat ttggccggcg gtttaagcct tgattgtcag tacttgggag aacataaaag
 10561 gtgtaataca aggtgcttta aatatcatac ttggcttgat taagtctctc tcaagtttat
 10621 tctgttggtg ttggcgagga gtttgggagc ccgtgtgat gattctttaa ggagcagttc
 10681 aattaaattg gaatttagtt caattatggg ttgtaggtaa aataactggg gttgttaggt
 10741 actttggcgg gttgctaaaa ggattgatat cagggaattt ggacgtaata agaagtatat
 10801 tcagtaaatc ttatcagca atttggatg caacaaaaag tatttttggg tttttattta
 10861 atagcgtaaa atcaattttc acaaatatga aaaattgggt atctaatact tggagcagta
 10921 tccgtacgaa tacaatagga aaagcgagc cattatttag tggcgtcaaa tcaaaattta
 10981 ctaatttatg gaatgcgacg aaagaaattt ttatgtaatt aagaaatttg atgtcaata
 11041 ttgggaattc cattaaagat aatacggtag gaattgcaag ccgtttatgg agtaaggtag
 11101 gtgggaattt cacaatatg cgcatgggt tgagttccat tatagataag attaaaagtc
 11161 atatcggcgg tatggtaagc gctattaaaa aaggacttaa caaattaatc gacggtttta
 11221 actgggtcgg tggtaagttg ggaatggata aaatacctaa gttacacact ggtacagagc
 11281 acacacatc tactacaaga tttagttaaga acggtagat tgacagtgac acattcgcta
 11341 cagttgggga taaggggagc ggaaatgggt caaatgggtt tagaaatgaa atgattgaat
 11401 tccctaacgg taaacgtgta atcacaccta atacagatag taccgcttat ttacctaaag
 11461 gctcaaaagt atacaacggt gcacaaactt attcaatgtt aaacggaacg cttccaagat
 11521 tagtttagg tactatgttg aaagatatga aatctgggtg atcatcgga ttaactgga
 11581 caaaagataa aataggttaa ggtaccaaat ggcttggcga taaagtggc gatgttttag
 11641 attttatgga aaatccagcg aaacttttaa attatatact tgaagctttt ggaattgatt
 11701 tcaattcttt aactaaagggt atgggaattg caggcgacat aacaaaagct gcatggctta

```

11761 agattaagaa aagtgtctact gattggataa aagaaaaattt agaagctatg ggcgggtggcg
11821 atttagtcgg cggaaatatta gaccttgaca aaattaatta tcattatgga cgtaccgcag
11881 cttataccgc tgcactgga agaccatttc atgaagggtgt cgattttcca tttgtatatac
11941 aagaagttag aacgccgagtg ggtggcagac ttacaagaat gccatttatg tctgggtgggt
12001 atggtaatta tgtaaaaatt actagtggcg ttatcgatat gctattttgc catttgaaaaa
12061 acttttagcaa atcaccacct agtggcacga tggtaaagcc cgggtgatgtt gttgggttaa
12121 ctggtaatac cggatttagt acaggaccac atttaccattt tgaatgagg agaaatggac
12181 gacattttga cctgaacca tatttaagga atgctaagaa aaaaggaaga ttaccaatag
12241 gtgggtggcg tgctacttct ggaagtggcg caacttatgc cagtcgagta atccgacaag
12301 cgcaaagrat tttaggtggt cgttataaag gtaaatggat tcatgaccaa atgatgcccgc
12361 ttgcacaaacg tgaagtaaac taccagtcga atgcagtgaa taactgggat ataatgctc
12421 aaagaggaga cccatcaaga ggattattcc aaatcatcgg ctcaactttt agagcaaacg
12481 ctaaaactgg atatactaac tttaataatc cagtacatca aggrtatctca gcaatgcagt
12541 acattgttag acgatatggt tggggtgggt ttaaacgtgc tgggtattac gcatatgcta
12601 cagggtgaaa agtttttgat ggttgggtata acttaggtga agacgggtcat ccagaatgga
12661 ttattccaac agatccagct cgtagaaatg atgcaatgaa gattttgcat tatgcagcag
12721 cagaagtaag agggaaaaaa cgcagtaaaa ataagcgtcc tagccaatta tcagacttaa
12781 acgggtttga tgatcctagc ttattattga aaatgattga acaacagcaa caacaaatag
12841 cttttactact gaaaaatagca caatctaacg atgtgattgc agataaagat tatcagccga
12901 ttattgacga atacgctttt gataaaaaag tgaacgcgtc tatagaaaaag cgagaaagggc
12961 aagaatcaac aaaagtaag tttagaaaaag gaggaattgc tattcaatga tagacactat
13021 taaagtgaac aacaaaacaa ttccctgggt gtatgtcgaa agagggtttg aaataccctc
13081 ttctaattat gttttaaata cagaaaaatgt agatggacgt tccgggtcta tatataaagg
13141 gcgttaggctt gaatcttata gttttgatat acctttgggt gtacgtaatg actatttatc
13201 tcacaacggc attaaaacac atgatgacgt cttagaatgaa ttagtaaaagt tttttaacta
13261 cgaggaacaa gttaaattac aattcaaatc taaagattgg tactggaacg cttatttcga
13321 aggaccaata aagctgcaca aagaatttac aatacctggt aagttcacta tcaaaagtagt
13381 actaacagac ccttacaat attcagtaac aggaataaaa aatactgcga tttcagacca
13441 agtttcagtt gtaaatagtg ggactgctga cactccttta attgttgaag cccgagcaat
13501 taaaccatct agttacttta tgattactaa aaatgatgaa gattatttta tgggtgggtga
13561 tgatgaggta accaaagaa ttaaggatta catgcctcct gtttatcata gtgagtttcg
13621 tgatttcaaa ggttgacta agatgattac tgaagatatt ccaagtaatg acttaggtgg
13681 taagggtggc ggtgactttg tgatattcaa tcttggcgaa ggatataaag caactaattt
13741 tccgtgatgca aaagggttggg ttggtgctgg cagcaaacga gggctcccta aagcgatgac
13801 agattttcaa attacctata aatgtattgt tgaacaaaaa ggtaaagggtg ccggaagaac
13861 agcacaacat atttatgata gtgatggtaa gttacttgct tctattgggt atgaaaaataa
13921 atatcatgat agaaaaatag gacatattgt tgttactgtg tataaccaaa aaggagacct
13981 caaaaaagata tacgactatc agaataaacc gataatgtat aacttggaca gaatcggtgt
14041 ttatatgccc ctcagaagag taggtaataa attttctatt aaaacttggg aatttgatca
14101 cattaaagac ccagatagac gtaaacctat tgatattggat gagaaagagt ggatagatgg
14161 cggtaagttt tatcagcgtc cagcttctat catagctgtc tatagtgcga agtataaccg
14221 ttataagttg atggagatga atgggttagg ttcattcaat acggagattc taccgaaacc
14281 gaaaggcgca agggatgtca ttatacaaaa aggtgattta gtaaaaaatag atatgcaagc
14341 aaaaagtggt gtcattcaatg aggaaccaat gttgagcgag aaatcgtttg gaagtaatta
14401 tttcaatggt gattctgggt acagtgaatt aatcatacaa cctgaaaaacg tctttgtatc
14461 gacggttaaa tggcaagata gatatttata gaaaggagat gagagtgtga tacatgtttt
14521 agattttaac gacaagatta tagatttctt ttctactgat gaccttctct tagttagagc
14581 gattcataaa cgtaatgtta atgacaattc agaaatgctt gaactgctca tatcatcaga
14641 aagagctgaa aagttccgtg aacgacatcg tttattata agggattcaa acaaaacatg
14701 gcgtgaattt attattaact gggttcaaga tacgatggac ggctacacag agatagaatg
14761 tatagcgtct tatcttgctg atataacaac agctaaaccg tatgcaccag gcaaatttga
14821 gaaaaagaca acttcagaag cattgaaaaga tgtgttgagc gatcacagggt gggaagtttc
14881 tgaacaaacc gaatacagtg gcttacgtac tacgtcatgg acttcttctc aaactagata
14941 tgaagtttta aagcaattat gtacaacctt taaaatgggt tttagatttt atattgagct
15001 tagcttctaat accgtcaaag gtagatatgt agtactcaaa aagaaaaaca gcttattcaa
15061 aggtaaagaa attgaatatg gtaaaagattt agtcgggtta actaggaaga ttgatattgc
15121 agaaatcaaa acagcattaa ttgctgtggg acctgaaaaat gacaaaggga agcgtttaga
15181 gctagtgttg acagatgacg aagcgcaaa gtaattcaac ctacctatgc gctatatttg
15241 ggggatatat gaaccacaat cagatgatca aaatatgaat gaaacacgat taagtctctt
15301 agccaaaaca gagttaataa aacgtaagtc ggcagttatg tcatatgaga ttacttctac
15361 tgatttggaa gttacgtatc cgcacgagat tatatcaatt ggcgatcacg tcagagtaaa
15421 acatagagat ttttaaccgc catttgtatgt agaggcagaa gttattgctg aagaatataa
15481 cataatttca gaaaatagca catatacatt cggtcacact aaagagttca aagaatcaga
15541 attacgagaa gagtttaaca agcgattgaa cataatacat caaaagttaa acgataatat
15601 tagcaatatc aacactatag ttaaagatgt ttagatgggt gaattagaat actttgaacg
15661 caaaaatacac aaaagtata caccgcccaga aaatccagtc aatgatattg tttggtatga
15721 tacaagtaac cctgatgtg ctgtcttgctg atgatatggg aatgggtcgat ggattgaagc
15781 aacaccaaat gatgtgaaa aattaggtgg tataacaaga gagaaagcgc tattcagtga

```


15841 attaacaat atttttatta atttatctat acaacacgct agtcttttgt cagaagctac
 15901 agaattactg aatagcgagt acttagtaga caatgatttg aaagcggact tacaagcaag
 15961 ttttagacgct gtgattgatg tttataatca aattaaaaat aatttagaat ctatgacacc
 16021 cgaaactgca acgattggct ggttggtaga tacacaagct ttatttcttg agtatagaaa
 16081 gaaattacaa gatgtttata cagatgtaga agatgtcaaa atcgccattt cagatagatt
 16141 taaattatta cagtcacaat acactgatga aaaatataaa gaagcgttgg aaataatagc
 16201 aacaaaattt ggtttaacgg tgaatgaaga tttgcagtta gtcggagaac ctaatgttgt
 16261 taaatcagct attgaagcag cttagagaatc cacaaaagaa caattacgtg actatgtaaa
 16321 aacatcggac tataaaacag acaagacggg tattgttgaa cgtttagata ctgctgaagc
 16381 tgagagaacg acttttaaag gtgaaatcaa agataaagtt acgttaaacy aatatcgaaa
 16441 cggattggaa gaacaaaaac aatatactga tgaccagtta agtgatttgt ccaataatcc
 16501 tgagattaaa gcaagttattg aacaagcaaa tcaagaagcg caagaagctt taaaatcata
 16561 cattgatgct caagatgatc ttaaaagaaa ggaatcgcaa gcgtatgctg atggtaaaat
 16621 ttccggaagaa gagcaacgcy ctatacaaga tgctcaagct aaacttgaag aggcaaaaaa
 16681 aaacgcagaa ctaaaaggcta gaaacgctga aaagaaaagct aatgcttata cagacaacaa
 16741 ggtcaaaagaa agcacagatg cacagaggaa aacattgact cgctatgggt ctcaaattat
 16801 acaaaatggt aaggaaatca aattaaagac tactaaagaa gagtttaatg caaccaatcg
 16861 tacactttca aatatattaa acgagattgt tcaaaatggt acagatggaa caacaatcag
 16921 atatgatgat aacggagtgg ctcaagcttt gaatgtgggg ccacgtggta ttagattaaa
 16981 tgctgataaa attgatatta acggtaatag agaaataaac ctctctatcc aaaatctgcy
 17041 agataaagta gataaaacgy atattgtcaa cagtcttaat ttatcaagag agggctctga
 17101 tatcaatgct aatagaattg gaattaaagg cggtagacat aacagatatg ttcaaatata
 17161 gaatgattct attgaactag gtggtattgt gcaacgtact tggagaggga aacgttcaac
 17221 agacgatatt ttacgcgac tgaagacggg tcacctaaaga tttagaataa acaccgctgg
 17281 cggttcactt tatatgtcac attttggtat ttcgacttat attgatggg aaggctgaaga
 17341 cggtaggtta ctctggtacga ttcaatgggt ggataaaact tacagtataa gtggcatgaa
 17401 tgggtataaca atcaattcct atgggtgggt cgttgcaact acgtcagata ataactgggt
 17461 tgcttcggag tcttacgctt catcgaatat caaaagcaaa caggcaccgg tgtatttata
 17521 tccaaacaca gacaaagtgc ctggattaaa ccgatttgca ttcacgctgt ctaatgcaga
 17581 taatgcttat tcgagtgacg gttatattat gtttgggtct gatgagaact atgattacgg
 17641 tgccgggtatc aggttttcta aagaaagaaa taaagggtct gttcaaatg ttatggagc
 17701 atatgaacaa ggtggagata caacaatcga agcagggtat ggcaaattta atatgctgaa
 17761 acgacgtgat ggttaatagg atattcatat acagagtaca gacctactgt ctgtaggttc
 17821 agatgatgca ggagatagga tagcttctaa ctcaatttat agacgtactt attcggccgc
 17881 agctaatttg catattactt ctgctggcac aattggggct tcgacatcag cgcgtataa
 17941 caagttatct atcgaaaatc aatataacga tagagatgaa caactggaac attcaaaagc
 18001 tattcttaac ttacctatta gaacgtggtt tgataaaagct gagtctgaaa ttttagctag
 18061 agagctgaga gaagatagaa aattatcgga agacacctat aaacttgata gatagctagg
 18121 tttgattgct gaagaggtgg agaatttagg attaaaagag ttgtgtcacgt atgatgacaa
 18181 aggaagaaatt gaaggatagc cgtatgatcg tctatggatt catcttatcc ctgttatcaa
 18241 agaacaacaa ctaagaatca agaaattgga ggagtcaaa aatgcaggat aacaacaag
 18301 gattacaagc taatcctgaa tatacaattc attatttatc acaggaatatt atgaggttaa
 18361 cacaagaaaa cgcgatgtta aaagcgtata tacaagaaaa taaagaaaaa caacaatgtg
 18421 ctgaggaaga gtaaatcctta gcactatttt tatacaaaaa ttaaggaggg tcattttaatt
 18481 atggcaaaag aaattatcaa caatacagaa aggtttattt tagtacaatc cgacaaagaa
 18541 ggtacagaac gtgtagtata tcaagatttc acaggaagtt ttacaacttc tgaatgggt
 18601 aacattgctc aagattttta atctgaagaa aacgctaaga aaattgcgga gacgttaaat
 18661 ttgttatatt aatttaactaa caaaaaacaa cgtgtgaaag tagttaaaga agtagttgaa
 18721 agatcagatt tatctccaga ggtaacagtt aacactgaaa cagtatgaaa agctatgagt
 18781 tagatactca tagtctttat tcttttagaa agcgggtgta ctgaattggg gtggttcaaa
 18841 aaacacgaac atgaatggcg catcagaagg ttagaagaga atgataaaac aatgctcagc
 18901 acaactcaacg aaattaaatt aggtcaaaaa acccaagagc aagttaacat taaattagat
 18961 aaaaccttag atgctattca aaaagaaaga gaaatagatg aaaagaataa gaaagaaat
 19021 gataagaaca tacgtgatat gaaaatgtgg gtgcttgggt tagttgggac aatatttggg
 19081 tcgctaatta tagcattatt gcgtatgctt atgggcatat aagagagggt attaccatgt
 19141 tcggattaaa ttttgagctc tcgctgtgga cgtgtttctg gtttggttaag tgaagtaat
 19201 agttaagagt cagtgtctcg gcactggctt tttatttttg ataaaaaggg caaacaatg
 19261 gatgcaaaag taataacaag atacatcgta ttgatcttag cattagttaa tcaattctta
 19321 gcgaacaag gtattagccc aattcccagta gacgatgaaa ctatatcatc aataactctt
 19381 actgtagtcg ctttatatac aacgtataaa gacaatccaa catctcaaga aggtaaatgg
 19441 gcaaatcaaa aattaaagaa atataaagct gaaaataagt atagaaaagc aacagggcaa
 19501 gcgccaatta aagaagtaat gacacctag aatatgaacg acacaaatga tttagggttag
 19561 gtggttgata tatgttaagt acaaaaaatc aagcagaaaa atggtttgac aattcattag
 19621 ggaaacaatt caaccagat ggttggtatg gatttcagtg ttatgattac gccaatatgt
 19681 tctttatggt agcgacaggc gaaaggctgc aagggttata tgcttataat atccggttg
 19741 atcaataaagc aaagattgaa aatatgggtc aaataattaa aaactatgac agctttttac
 19801 cgcaaaagtt ggatattgct gttttcccg tcaagtatgg tggcggagct ggacacgttg
 19861 aaattgttga gagcgcaaat ttaaatactt tcacatcatt tggcctaaaac tggaaacgga

19921 aagggttgac taatggcggt gcgcaacctg gttgggggtcc tgaacctgtg acaagacatg
19981 ttcattacta tgacaatcca atgcatctta ttaggttaaa cttccctaac aacttaagcg
20041 ttggcaataa agctaagggt attattaagc aagcgactac aaaaaaagag gcagtaatta
20101 aacctaaaaa aattatgctt gtagccggtc atggttataa cgatcctgga gcagtaggaa
20161 acggaacaaa cgaacgcgat tttatcacga aatatataac gcctaataac gctaagtatt
20221 taagacatgc aggcacatgaa gttgcattat acggtggctc aagtcaatca caagatatgt
20281 atcaagatac tgcatacggg gttaatgtag gcaataaaaa agattatggc ttatattggg
20341 ttaaatcaca ggggtatgac attgttctag aaatacattt agacgcagca ggagaaagcg
20401 caagtgggtg gcattgtatt atctcaagtc aattcaatgc agatactatt gataaaagta
20461 tacaagatgt tattaataat aacttaggac aaataagagg tgtgacacct cgtaagtatt
20521 tactaaatgt taatgtatca gcagaaataa atataaatta tctgttatct gaattagggt
20581 ttattactaa taaaatgat atggattgga ttaagaaaaa ctatgacttg tattctaaat
20641 taatagccgg tgcgattcat ggtaagccta taggtgggtt ggtagctggg aatgttaaaa
20701 catcagctaa aaacaaaaaa aatccaccag tgcacgagg ttatacactc gataagaata
20761 atgctcccta taaaaaagaa caaggcaatt acacagtagc taatgttaaa ggtaataatg
20821 taagagacgg ttattcaact aattcaagaa ttacaggggt attacccaac aacacaacaa
20881 ttacgtatga cgggtacatg tgtattaatg gttatagatg gattacttat attgctaata
20941 gtggacaacg tctgtatata gcgacaggag aggtagacaa ggcaggtaat agaataagta
21001 gttcttggtaa gtttagcacg atttagtatt tacttagaat aaaaattttg ctacattaat
21061 tatagggaat cttacagtta ttaataaact atttggatgg atgttaatat tctatcacac
21121 ttttttaaat tttctctcaag atttaaatgt agataacagg caggtaacttc ggtacttgcc
21181 ttttttttta tgttatagct agccttcggg ctatgttttt gttatgatgt gttacacatg
21241 catcaactat ttacatctat ccttgttctc ccaagcatgt cactggatgt tttttcttgc
21301 gatagataga atagtcttca tactactccc cgtagtatat atgactttag cttcccgta
21361 taacagttta cggggtgctt ttatgttata attgctttta tatagtagga gtgaactata
21421 tagccgggca gagggcatgt atctgactgt tgggtccaca ggagacatct tcttctgtat
21481 cactcgatgc atatatctta acaacataga aatgttatcat tctgtataac cgtatcttaa
21541 tctgatacggg tatattttat cccctacaac caacaaaacc acagatccta ttaatttagg
21601 attgtggtta ttttttgctt ttttttgggg caaaaaaagg gcagattatt tgaaaaaggg
21661 caaacgcttg tggaaaagct aaaagggttaa aaatgacaaa aaccttgata caacagtgtt
21721 ttgggacgct cgtgtacggt agagaatgac cgytttacc aacatacaag gtgggattaa
21781 cttgtgttaa aaagccttta atatcagttg ttacaaagga tttgtagcgt ctttaaaaat
21841 aaaaaagggc agaaaaaggg cagataacct ttatgtacac agtttttcta atttttgctc
21901 taactctctg tccattttct ctgttacatg tgtatacacc ttatagtcg tttttctatc
21961 tgtatgtcct actcttttca taattgcttt taacgatata tttatttcg ccaataaact
22021 tatgtgtgta tgccttagtg tgtgagtagt aactttttta tttatattta atgattctgc
22081 agctgaggac aatcgtttgt ttatcctact gccttgcata ggatttctct ggcaagttgt
22141 gaataataac cctctatcaa catagcttgg ttcccatgtg tgcactcttt tattttctaa
22201 cattattttt ttcaatacat ttgctatcct tgaattgatg gcgatttttc tttctgaacc
22261 tgcggtctta gtatgatctt tgtgaccaa tccagcatta catttgattc tgtgaatagt
22321 gccatttaata gcgatcgttt tatttttgag gtcaacatct ttaacttgga gagctaataa
22381 ctcacctatg cgcatacctg ttaaaagctt aacttttaca gcccagcaa ctaaaatagc
22441 agctctatgc tgcattgtat tatcgttcag tataaaatcg cgtatctgta ttacctgttc
22501 catctctaaa tagttatata tttctcgttc tttttttct atatcttcta tctgttact
22561 cttctttggt agtgtgacgc tatttaatat gtgttctgtt ggataattgt aaaaattaac
22621 ggcgtattta atagcttctt tcatatgtcc aagttgacgc tttacctgat ttgcagaata
22681 tacgtttgat aatttgttaa taaatgtttg catgtacttt gtatcaattt tgtttaaaag
22741 taaattttga gaactgttct ttttgatgtt ttgtattctt gttttcaaat tatcaagcgt
22801 cgttacttta aagccagatg ttttttatag atattcaagc cattcatcta ataacgcgtg
22861 aaaagtcaaa gtttttaatt cgttgacga cttgttgttt agtttttctt ttttttttc
22921 ttctaacaag aacattgcct ctttttgcca ttgctttgta ttttttatca agacaacact
22981 tacacgcttc catttatctg tatacggatc tttgtatttc tctgtatgc tttacttctg
23041 tttctgttct ttatttttaa atttttcaaa ccacatttta catccctcct caaaaattggc
23101 aaaaaataat aagggttaggc gggctaccca tgaataattgt aaaaaaaaag acgcttctgt
23161 aaaaatacaga cggcacttat aattataaga ttacatgggt aattacaaa aatggtaacg
23221 aatatatacg tgttttaaaag gataaacctt taatatatta aaattatata atcttatata
23281 agggatctgc aatatattat tattaattct atttatcagt aacataatac ccgaagaate
23341 tattactgga tttttaattt ttggggtaa aacttttctt atgcgaact tactaatcgg
23401 ctggaagaa tttatgcaag cgttaactatt accttttaatt ttttttacct tatcaattgc
23461 tgatactatg ttattaatgt ttctgtcaat ttatttttaa tttctaaact tttctaaact
23521 atcagatata aattcaataa aataatcttt agtgatgaat tctgtgtgtt ttttttggt
23581 ttttttatcg aaaacttctt ttaatatagc tgaattattt tgcgcgctaa ttaaatttaa
23641 aacaactctt aaataactat cccatttcaa atcaaaattc atctttaaact aacttttgtt
23701 tttcttagag gataaggaa taacatttac tatatctctc gttatagaat catttttatt
23761 catcactatt gcaagtggtg aattagaaaa tttctttatta acgtttatata cgaatcttac
23821 aaaaactatt tctccttgtt taaactttgg ataaaaacct ttatggtttt tttcaccttc
23881 aaatctcttg agtaaatagt gaatatctga atccaacttt ttaaattttg gatttccaga
23941 agtttttaatt ttattaatgc gtttttctat attatgcgtc atcattttctc ctttattctc

24001 gctcacactc tcaccacccat tcaacgtcta cacttgragg cgttttttga ttagtaaaat
24062 cataatgaat cttcttttgg taacttatcg ccatctatrt tttgtgaaat aaattccaag
24121 tattttacgag cattatgtga cgataaatct ttaggttaact cataagtga tgggtgatta
24181 ccactagtta aaacttcata tactatagtt tcttttttta ttttgcaatt agtrattttc
24241 attataaact ccttttaaac acrgctgaaa tagacgtctt tttcaataa gcargattaa
24301 tactttaatt ctttaatcca catatattta aaagttaggt agtaggtaat aaatacaaga
24361 crraaagtta agattgcttt tttcatgtca atttctcctt tgtttatatt tatattaaag
24421 cgctaaatat acgttattta tcacaataca actttgccc tttactttaat atcactaaac
24481 gaagcgactt tgatatcatc atacttcgga tttagagata ccaaattaat atagtcttcg
24541 catatatcta cagcgttgat aagacttact ccatcctaata caacgagtcg aattgtacca
24601 cctttaatag aatcttcttt ctttaataaaa gcgtatgttc cttgttttaa cataggttcc
24661 attgaatcac catttaactaa aatacaaaaa tcagcatttg atggcgtttc gtcttcttta
24721 aaaaataact cttcatgcaa tatgtcatca tataattctt ctcctatgcc agcaccaggt
24781 gcaccacatg caatatatcg tactagttta gactccttat attcatctat agaagtgact
24841 ttatttctgt catctaattg ctcatgttga tagttaagta cgttttcttg gcggggaggt
24901 gtgagttgag aaaatatggt attgattttt gacattatcg tttcatcttg acgttcttcg
24961 tcagggaactc gataagaatc tacatcatat cccataagcc acgttccacc gacatttaaa
25021 gtttttagata ataagaataa tttatgttgg tctggagaag accttccatt aacatactgg
25081 gataagtgcac tttttgacat ttttaatttc aattcttttt gaaaggggtt cgacttttct
25141 agaatactta cttgacgcaa gttcctatct ttcataattt gttttaatct ttcagaagtg
25201 ttttgcattg gtaatgcctc cttgaaattc attatatagg aagggaataa aaaatcaata
25261 caaaagtcca acttttttaa ctttttgtgt tgacattgtt caaaattggg gttatagtta
25321 ttatagttcca aatgtttgaa ctttaggggt gattatttga atactaatat aacttttgat
25381 ttttcgttat tgaacggtta gatagtcgaa gtgtactcga cacaatttaa ctttgctata
25441 gcttttaggtg tatcagaaag aactttgtct ttgaagtga acaacaaagt accatggaaa
25501 acaacagaca ttattaaagc ttgtaagtta ttgggaatac ctataaaaga tgttcacaaa
25561 tattttttta aacagaaagt tcaaatgttt gaacttaata agtaaaaggag gcataacaca
25621 tgcaagaacg agaaaagggt aataaaaagta acacatcttc aaatgaagca tcaaacctt
25681 ttaggacaaa ttgaagctta cgacaaaacg cttaaagaaa taaagtacac tcgagacctt
25741 tacaacaacac acctaagcat gaacaacgaa gacgcattcg ctggtttgg aatggttagg
25801 gatgaataa ctaaaaagct acgaagtgtc atcaaaagat tccaaaaggt agtgaagcg
25861 ttagacaagc ttaacggtgt tgaagcgcat aacaaggtta ctgatttaac agagtggcgg
25921 aaagtgaatc agtaacattc acttcttaat ataaccacgc ttatcaacat ccacattgag
25981 cagatgtcag cgagagctgg cgatgatag agccgcgttt aatatcattc gatagtcatt
26041 gcgataaccg tctgctgaat gtgggtgttg aggaaaaagg aggatactca aatgcaagca
26101 ttacaaacat ttaattttta agagctacca gtaagaacag tagaaattga aaacgaacct
26161 tattttgtag gaaaaatat tgcgtgagat ttaggatag caagatcaaa caatgccatt
26221 agaaatcatg ttgatagcga ggacaagctg acgcaccaat ttagtgcatc aggtcaaaac
26281 agaaatcatg tcattatcaa cgaatcagga ttatcacgct taatcttcga tgcctcttaa
26341 caaagcaaaa acgaaaaaat tagagaaacc gctagaaat tcaaacgctg ggtaacatca
26401 gatgtcttac cagctattcg caaacacggt atatacgcaa cagacaatgt aattgaacaa
26461 acattaaaag atccagacta catcattaca gtgtcgactg agtataagaa agaaaaagag
26521 caaaacttac ttttacaaca gcaagtagaa gtttaacaac caaaagtatt attcgtgac
26581 ttcgttagctg gtagtgataa ttcaatctt gttggagaac tagcgaaaaa acttaaacaa
26641 aacggtgttg atataggaca aaacagattg ttcaaatggt taagaaataa tggatatctc
26701 attaaaaaga gtggagaag ttataactta ccaactcaa agagtatgga tctaaaaatc
26761 ttggatctca aaaaacgaat aattataat ccagatggtt caagttaagt atcacgtaca
26821 ccaaaagtta caggcaagg acaacaatc tttgttaata agtttttagg agaaaaacaa
26881 acatcttaaa aggaggaaca caatggaaca aatcacatta accaaagaag agttgaaga
26941 aattcacaga aaagaagtta gagaggttat aaatggcaag aaaccaatca gttcaggttc
27001 aattttcagt aaagtaagaa tcaataatga cgatttagaa gaaatcaata aaaaactcaa
27061 tttcgcaaaa gattttgtcg taggaagatt gaggaagctc aatcatccga ttcggttaa
27121 aaagtatcag catggcttcg aatcaattca tcaaaaagct tatgtacaag atgttcatga
27181 ccatattaga aaattaacat tatcaatttt tggagtgaac ctttaattcag acttgagtga
27241 aagtgaatc aacctagcag caaaagttaa tcgagaaatc aaaaactatt atttatacat
27301 ctatgaaaag agagtttcag aattaaactat cgatgatttc gaataaagga ggaacaacaa
27361 atgttacaaa aatttagaat tgcgaagaa aaaaaataat taaaactcaa attactcaag
27421 catgctagtt actgtttaga aagaaacaac aacctgaaac tgttgcgagc agttgcagag
27481 ttgttgaaaa aggttagcta aattcaacgg taaggatttg cctcgctcc acacttagag
27541 tttagatcc acaaacaca taagttttag tagggtctag aaaaaatgtt tgcatttctc
27601 cttttgtaac agtttcaatt ccttcataat cttgaaaaac aattttcttt aaatccgaaa
27661 catgtttttt tgaaccatcc tttaaagtaa ctagaagttt catacttatc acctccttag
27721 gttgataaca acattatata cgaaaggagc ataaacaata tgcaagcatt acaaacaaat
27781 tcgaacatcg gagaaatgct caatattcaa gaaaaagaaa atggagaat cgcaatcagc
27841 ggtcgagaac ttcatcaagc attagaagtt aagacagcat ataaagattg gtttccaaga
27901 atgcttaaat acggatttga agaaaaatca gattacacag ctatcgctca aaaaagagca
27961 acagctcaag gcaatatgac tcactatatt gaccacgcac tcacactaga cactgcaaaa
28021 gaaatcgcaa tgattcaacg tagtgaacct ggcaaacgtg caagacataa tttcatccaa

28081 gttgaaaaag catggaacag cccagaaaatg attatgcaac gtgcttttaa aattgctaac
28141 aacacaatca atcaattaga aacaaagatt gcacgtgaca aaccaaaaaat tgtatttgca
28201 gatgcagtag ctactactaa gacatcaatt ttagtggag agttagcaaa gatcattaaa
28261 caaaacggta taaacatcgg gcaacgcaga ttgtttgagt ggttacgtca aaacggattc
28321 cttattaaac gcaagggtgt ggattataac atgcctacac agtattcaat ggaacgtgag
28381 ttattcgaat ttaaagaaac atcaatcaca cattcggacg gtcacacatc aattagtaag
28441 acgcaaaaag taacaggtaa aggacaacaa tactttgtta acaagttttt aggagaaaaa
28501 caaacaactt aataggagga attacaaatg aacgcactat acaaaacaac cctcctcatc
28561 acaatggcag ttgtgacgtg gaagggtttg aagattgaga agcacactag aaaacctgtg
28621 attagtagca gggcggttgag tgactatcta aacaacaaat ctttaaccat accgaaagat
28681 gctgaaaaat ctactgaatc tgctcgtcgc cttttgaagt tcgccaacaa aactattagc
28741 aaataacaac attatacacg aaaggaaaga tagaaatgcc aaaaatcata gtaccacca
28801 caccagaaaa cacatataga ggcgaagaaa aatttgtgaa aaagtatac gcaaacctta
28861 cacaatccca tcaattgttt ggagtatgta gaagtacagt atacaactgg ttgaaatatt
28921 accgcaaaag taatttaggt gtgaaaaatt tatacattga ttattacca acaggcactc
28981 tgattaatat ttctaaattg gaagagtatt tgatcagaaa gcataaaaaa tgggtattag
29041 aggatattaa atgagcaaca ttataaaaag ctacctagta gcagtattat gcttcacagt
29101 cttagcgatt gtacttatgc cgtttctata cttcactaca gcattggtcaa ttgctggatt
29161 cgcaagtatc gcaacattca tgtactacaa agaattgctt ttcaaagaat aaaaaaactg
29221 ctacttgttg gagcaagtaa cagtatcaaa cacttaagaa aaaattcatg ttcaatataa
29281 aacgaaaaac ggagggaagtc aagatgtatt acgaaatagg cgaatcata cgcaaaaaa
29341 ttcatgttaa cggattcgtt tttaagctat tcattttaaa aggtcatatg ggcataatca
29401 tacaagttaa agatatgaac aacgtaccaa ttaaacatgc ttatgtcgtg gatgagaatg
29461 acttagatat ggcattcagac ttattttaacc aagcaataga tgaatggatt gaagagaaca
29521 cagacgaaca ggacagacta attaaacttag tcatgaaatg gtaggagggtc gctatgaagc
29581 agactgtaac ttatatcatt cgtcataggg atatgccaat ttatataact acaaaacca
29641 ctgataacaa ttcatgattt agttactcca caaatagaaa tagagctagg gaggtttaacg
29701 gtatggaaag agcgagtatc aatatggatt atcacaaagc aatcaagaaa acagtgcacg
29761 aaactattga gtacaggag gtgaacatg actgaggaaa acaagaagacc acaagaaaaa
29821 gtaagcatac tcaaaaaact aaagataaat aatatcgctg agaaaaataa aaggaaattc
29881 tataaatttg cagtatacgg aaaaattggc tcaggaaaaa ccacggtttgc tacaagagat
29941 aaagacgctt tcgtcattga cattaacgaa ggtggaacaa cggttactga cgaaggatca
30001 gacgtagaaa tcgagaacta tcaacacttt gtttatgttg taaatttttt acctcaaat
30061 ttacaggaga tgagagaaaa cggacaagaa atcaatgttg tagttattga aactattcaa
30121 aaacttagag atatgacatt gaattgatgt atgaaaaata agtctaaaaa accaacgttt
30181 aatgattcgg gagaagttgc tgaacgaatt gtcagtatgt acagattaat agggaaaact
30241 caagaagaat acaaatccca ctttgttatt acaggctcat aagggtatcaa caaagataaa
30301 gatgatgaag gtacgactat caaccctact atcactattg aagcgcaaga acaaatataa
30361 aaagctatta cttctcaaa gtatgtgtta gctagggcaa tgattgaaga atttgatgat
30421 aacggagaaa agaaagctag atatatctta aacgctgaac cttctaatac gtttgaaca
30481 aagatttagc attcaccttc aataacaatt aacaataaga aatttgcaaa tcttagcatt
30541 acggacgtag tagaagcaat tagaaatgga aactaaaaat taattaaaag gacggtattt
30601 aattatgaaa atcacaggac aagcgcaatt tactaaagaa acaaatcaag aaaagtttta
30661 taacgggtca cggagggttc aagctggaga attcacagtg aaagttaaaa atattgaatt
30721 caatgataga gaaaatagat atttcacaat cgtatttgaa aatgatgaag gcaaaacaata
30781 taaacataat caatttgtag cgccgtataa atatgatttc caagaaaaac aattgattga
30841 attagttact cgattaggta ttaagttaaa tcttcctagc tttagatttg ataccaatga
30901 tcttattggt aagttttgtc acttggtatt gaaatggaaa ttcaatgaag atgaaggtaa
30961 gtatttttac gatttttcat ttattaaacc ttacaaaaag ggcgatgatg ttgttaacaa
31021 acctattccg aagacagata agcaaaaagc tgaagaaaat aacggggcac aacaacaaac
31081 aaactgtctt caacaaagca atccatttga aagcagtggc caatttggat atgacgacca
31141 agatttagcg ttttaagggt tgggtttaa atcaatcatt acaagatacc agaaagataa
31201 cgacggtagc tattccgtcg ttgctactgg tgttgaaact gaacaaagtc acctgactt
31261 actagaaaaa ggaatatccac taaaagcaga agtagagggt ccggacaata aaaaactatc
31321 tatagaacaa cgcaaaaaaa tattcgcaat gtgtagagat atagaacttc actggggcga
31381 accagtagaa tcaactagaa aattattaca aacagaattg gaaattatga aagggtatga
31441 agaaatcagt ctgcgcgact gttctatgaa agttgcaagg gaggtaatag aactgattat
31501 agcgtcttag ttctcatcat aaatacctat gagtgtagaa acgagtaagt tgcataagcga
31561 agataaaagc ttatttatatt gggctacaat caaccgcaac tgtgtaatat gcggaagcc
31621 tcacgcagac ctggcacatt atgaagcagt cggcagaggc atgaacagaa acaaaatgaa
31681 ccactatgat aaacatgtat tagcgttatg tcgcgaacat cacaacgagc aacatgcgat
31741 tggcgttaa gttgttgatg ataaatacca cttgcatgac ctgtggaataa aagttgatga
31801 gaggctcaat aaaatgttga aaggagagaa aaaggaaatga atagactaag aataataaaa
31861 atagcactcc taatcgtcat cttggcggaa gagattagaa atgctatgca tgcgtgtaaaa
31921 gtggagaaaa tttttaaact tccgttttag taatacaggt ttttcaaaa gctttaccat
31981 agcgggacaa actaattgag ccttttttga tgcctattac ccagggggctg taatgtaact
32041 ttaatacttc aaattcaatg ccagaaaggt tacttattgt ttctagggtg tgccttgact
32101 ttaacattct ttttaacaaat tctaattccg aaacaaatct ttgtttttct ataattctat

```

32161 taaagtgtatt taaaaactga ggagcacaata acttattata aattccctttt ttgtttaagt
32221 aagacatgtc aaaagtttca tttaaaaccc ctaaccttac taggttatta attgaaattt
32281 cgggttgattc tatatctaac ggagagtcctt ttattaacgt gtccgatata ttcataccgt
32341 catctcttgg gtttaaaacc gctctatatt taacggcagg atgtacttcg tgattcctta
32401 aatgttttaa aagaatagca tcatttgggg ataattgttt aattatttca acaaatgaat
32461 ggtgggttaa tgagtctttt ctgtcatcca tagatgargc tattagtttt gcgaacatat
32521 tacttaaaagt tttttcacta atgtaaaact ttgaagcttc tagagcagga cctagaagag
32581 aaaattgtgg ttcttgtaaa ttatttttag gtacagaaga tatttctttt ttaattgtt
32641 ctttgaattt ttcaaatctt acttctcttt gataaataac tttatccaca taaagggtgga
32701 atttcccaaa gacaagtctc caagtttttag agaattgttc tacaggccct ttgatgctc
32761 cttcaataat tttatcaata cctttacctt aaataggatc cataattatt caccctcaat
32821 ctaacgcaat agcgataata aaattatacc agaaaggaga atcaacatga ctgaccaacc
32881 aagttactac tcaataatta cagcaaatgt cagatacgat aaccgactta ctgacagcga
32941 aaagttaact tttgcagaaa taacatcttt aagtaacaaa tacggatact gcacagcaag
33001 taatgggttct tttgcaactt tatacaacgt tgttaaggaa actatatctc gragaatttc
33061 gaaccttacc aactttgggt atctaaaaat cgaattatc aaagaaggta atgaagttaa
33121 acaaaaggag atgtacctct tgacgcaaac gtcaatacct attgacgcaa aaatcaatac
33181 ccctattgat aattctgtca atacccttat tgacgcaaat gtcaaaagaga atattacaag
33241 tattaataat acaagtaata acaatataaa tagaatagat atattgtcgg gcaacccgac
33301 agcatctctc ataccctata aagaaattat cgactactta acaaaaaaag cgggcaagca
33361 ttttaaacac aatacagcta aaacaaaaga ttttattaaa gcaagatgga atcaagattt
33421 taggtttgag gatttttaaa aggtgattga tatcaaaaca gctgagctgc taaacacgga
33481 tagcgtataa taccttagac cagaaacact ttttggcagt aaattttgag ggtacctcaa
33541 tcaaaaaata caaccaactg gcacggatca attggaacgc atgaagtacg acgaaagtta
33601 ttgggattag ggggatatta tgaaccact attcagcgaa agataaaacg aaagcttgaa
33661 aaaatatcaa cctactcatg tcgaaaaagg attgaaatgt gagagatgtg gaagtgaata
33721 cgacttatat aagtttgctc ctactaaaaa acaccgaat ggttacgagt ataaagacgg
33781 ttgcaaatgt gaaatctatg aggaatataa gcgaaacaag caacggaaga taaacaacat
33841 attcaatcaa tcaaacgtta atccgtcttt aagagatgca acagtcaaaa actacaagcc
33901 acaaaatgaa aaacaagtac acgctaataa aacagcaata gactacgtac aaggcttctc
33961 tacaaaaaaa ccaaaatcat taattattga aggttcatac ggaactggta aaagccacct
34021 agcatacgct atcgcaaaag cagtcaaaagc taaagggcat acggttgctt ttatgcacat
34081 accaatgttg atggatcgta tcaaaaggac atacaacaaa aatgcagtag agactacaga
34141 cgagctagtct agattgctaa gtgatattga tttacttgta ctagatgata tgggtgtaga
34201 aaacacagag cacactttaa ataaaacttt cagcattgtt gataacagag taggtaaaaa
34261 caacatcttt acaactaact ttagtataa agaactaaat caaaatatga actggcaacg
34321 tataaatctg agaattgaaa aaagagcaag aaaagtgaag gtaatcggag acgatttcag
34381 ggagcgagat gcatggtaac caaagaattt ttaaaaacta aacttgagtg ttcagatatg
34441 tacgctcaga aactcataga tgaggcacag ggcgatgaaa atagggttga cgacctattt
34501 atccaaaaac ttgcagaacg tcatacacgc cccgctatcg tcgaatatta aggagtgtta
34561 aaaaatgccga aagaaaaata ttacttatac gcagaaagatg gcacagaaga tattagggtc
34621 atcaagrata aagacaacgt aaatgaggtt tattcgctca caggagccca tttcagcgac
34681 gaaaagaaaa ttatgactga tagtgacctt aaacgattca aaggcgctca cgggcttcta
34741 tatgagcaag aattaggttt acaagcaacg atatttgata tttagaggtg gacgatgagt
34801 aaatacaacg ctaagaaagt tgagtacaaa ggaattgtat ttgatagcaa agtagaggtg
34861 gaattattacc aatattttaga aagtaatatg aatggcacta attatgatca tatcgaaata
34921 caaccgaaat tcgaattatt accaaaacta gataaacaac gaaagattga atatttgca
34981 gacttcgctg tatatctcga tggcaaaact attgaaatta tcgacattaa aggtatgcca
35041 accgaagtag caaaacttaa agctaagatt ttcagacata aatcacagaaa cataaaactc
35101 aattggatat gtaaaagcgc taagtataca ggtaaaacat ggattacgta cgaggaatta
35161 attaaagcaa gacgagaacg caaaagagaa atgaagtgat ctaatgcaac aacaagcata
35221 tataaatgca acgattgata taaggatacc tacagaaggt gaatatcagc attttgatga
35281 tgtggataaaa gaaaaagaa gctcggcaga ttactttat atcaatctctg acgaaatact
35341 agagtatgac aattttaaaa ttagaacgt aaatgtagag gtggaataaa tgggcagtgt
35401 tgtaatcatt aataataaac catataaatt taacaatttt gaaaaaagaa ataattggca
35461 agcgtgggat aaatgctgga attgtttcta aacgtgttag aggttggttg gagttttcag
35521 aagcttttaga cgcgccttat ggcattgcacc taaaagaata tagagaaatg aaacaaatgg
35581 aaaagattaa acaagcgaga ctcgaaacgt aattggaaag agagcgaaag aaagaggtcg
35641 agctacgtaa gaagaagcca catttggtta atgtacctca aaaacattca cgtgatccgt
35701 actgggttca tgtcacttat aaccaaatgt tcaagaaatg gagtgaagca taatgagcat
35761 aatcagtaac agaaaagtag atatgaacaa aacgcaagac aacgttaagc aacctgcgca
35821 ttacacatac ggcgacattg aaattataga ttttattgaa caagttacgg cacagtaccc
35881 accacaatta gcattcgcaa taggtaatgc aattaaatag ttgtctagag caccgttaaa
35941 gaatggctcat gaggatttag caaaggcgaa gttttacgtc gatagagtat ttgacttgtg
36001 ggagtgatga ccatgacaga tagcggacgt aaagaatact taaaacattt ttcggctctc
36061 aagagatatc tgtatcagga taacgaacga gtggcacata tccatgtagt aaatggcact
36121 tattactttc acggtcatat cgtgccaggt tggcaagggt tgaaaaagac atttgatata
36181 gcggaagagc ttgaaacata tataagcaa agtgattttg aatatgagga acagaagcaa

```

```

36241 ctaactttat tttaaaagg cggaaacaat gaaaatcaaa attgaaaaag aaatgaattt
36301 acctgaactt atccaatggg cttgggataa ccccaagtta tcaggtataa aaagattcta
36361 ttc aaatgat gttgagcgca actgttttgt gacttttcat gttgatagca tcttatgtaa
36421 tgtgactgga tatgtatcaa ttaacgataa atttactgtt caagaggaga tataacaatg
36481 aaaatcaaaag ttaaaaaaga aatgagatta gatgaattaa ttaaatgggc gcgagaaaat
36541 ccggatctat cacaaggaaa aatatttttt tcaacaggat ttagtgatgg attcgttcgt
36601 tttcatccaa atacaaataa gtgttcgacg tcaagtttta ttccaattga tatcccttc
36661 atagttgata ttgaaaaaga agtaacggaa gagactaagg ttgatagggt gattgaatta
36721 ttcgagattc aagaaggaga ctataactct acactatatg agaacactag tataaaagaa
36781 tgtttatatg gcagatgtgt gcctaccaa gcatcttaca tcttaaacga tgacctaaat
36841 atgacgttaa tctggaaaga tggggagttg ctagtatgat gttgaaattt aaagcttggg
36901 ataaagataa aaaagtattg agtattattg acgaaatcga ttttaattag ggttacattt
36961 tgatttcaac aggttataaa agtttcaatg aagtaaaact attacaatac acaggattta
37021 aagatgtgca cgggtgtggag atttatgaag gggatattgt tcaagattgt tatccgagag
37081 aagtaagttt tatcgagttt aaagaaggag ctttttatat aacttttagc aatgtaactg
37141 aattactaag tgaaaatgac gatattattg aaattgttgg aaatatttt gaaaatgaga
37201 tgctattgga ggttatgaga tgacgttcac cttatcagat gaacaatata aaaatctttg
37261 tactaactct aacaagttat tagataaaat tcacaaagca ttaaaagatc gtgaagagta
37321 caagaagcaa cgagatgagc ttattgggga tatagcgaag ttacgagatt gtaacaaaga
37381 tctagagaaag aaagcaagcg catgggatag gtattgcaag agcgttgaag aagatttaat
37441 aaacgaattc ggttaacgat atgaaagagt taaattcggga atggaattaa acaataaaat
37501 ttttatggag gatgacacaa atgaataatc gcgaaaaaat cgaacagtcg gtatttagtg
37561 ctagtgcgta taacggtaat gacacagagg ggttgctaaa agagattgag gacgtgtata
37621 agaaagcgca agcgtttgat gaaatacttg agggaatgac aaatgctatt caacattcag
37681 ttaaaagagg tattgaaact gatgaagcag tagggattat ggcagggtcaa gttgtctata
37741 aatatgagga ggaataggaa aatgactaac acattacaag taaaactatt atcaaaaaat
37801 gctagaatgc cgaacgaaa tcataagacg gatgcagggt atgacatatt ctcagctgaa
37861 actgtcgtac tcgaaccaca agaaaagca gtgacaaa cagatgtagc tgtgagtata
37921 ccagagggtc atgtcggact attaaactagt cgtagtgttg taagtagtaa aacgtattta
37981 gtgattgaaa caggcaagat agacgcggga tatcatggca atttagggat taatatcaag
38041 aatgatgaag aacgtgatgg aatacccttt ttatatgatg atatagacgc tgaattagaa
38101 gatggattaa taagcatttt agatataaaa ggtaactatg tacaagatgg aagaggcata
38161 agaagagttt accaaatcaa caaaggcgat aaactagctc aattggttat cgtgcctata
38221 tggacaccgg aactaaagca agtggaggaa ttcgaaagtg tttcagaacg tggagcaaaa
38281 ggcttcggaa gtacgagagt gtaaagacat cttagatcga gttaaaggag ttttggggaa
38341 gtgacgcaat acttagtcac aacattcaaa gattcaacag gacgaccaca tgaacatatt
38401 actgtggcta gagataatca gacgtttaca gttattgagg cagagagtaa agaagaagcg
38461 aaagagaagt acgaggcaca agttaaaaga gatgcagtta ttaaagtggg tcagttgtat
38521 gaaaaatataa gggagtgttg gaaatgacgg atgttaaaat taaaactatt tcaggtggag
38581 tttattttgt aaaaacagct gaaccttttg aaaaatattg tgaagaatg acgagtttta
38641 atgtttatat ttacgcaagt actataatca agaaaccaac gtatatataa acagatacga
38701 ttgaatcaat cacacttatt gaggagcatg ggaatgaat cagctgagaa ttttattaca
38761 tgacggtagt agtttgatat tacatgaaga tgaattattt aacgaatatg tatttgtttt
38821 ggacaatttt agaaatgat atgactattt aacgatagaa aaagattatg gcagagaact
38881 tgtattgaac aaaggttata tagttgggat caatgttgag gaggcagatg atgattaaca
38941 tacctaaaat gaaattcccg aaaaagtaca ctgaaataat caaaaaatat aaaaaaaag
39001 cacctgaaga aaaggctaa attgaaatg attttattaa agaaattaaa gataaagaca
39061 gtgaatttta cagtcctacg atggctaata tgaatgaata tgaattaaag gctatgttaa
39121 gaatgatgcc tagtttaatt gatactggag atgacaatga tgattaaaaa acttaaaaat
39181 atggatgggt tcgacatctt tattgttggg atactgtcat tattcgggat attcgcattg
39241 cracttggtt tcacattgcc tatctataca gtggctagtt accaacacaa agaattacat
39301 caaggaaacta ttacagataa atatacaag agacaagata aagaagacaa gttctatatt
39361 gtattagaca acaaacagat cattgaaaat tccgacttat tattcaaaaa gaaatttgat
39421 agcgcagata tacaagctag gttaaaagta ggcgataagg tagaagtaa aacaatcggg
39481 tatagaatcac acttttttaa tttatatccg gtcttatacg aagtaaagaa ggtagataaa
39541 caatgattaa acaaatacta agactattat tcttactagc aatgtatgag ttaggtaagt
39601 atgtaactga gcaagtgtat atttatgata cggctaatag tgatgtagag gcgccgagtg
39661 attacgtctt tcgagcggag gtgagtgaat aatgagaata tttatttatg atttgatcgt
39721 tttgtctgtt gcttctctaa tatccatata tattattgat gatggagtga taataaatgc
39781 attaggaatt tttggatagt ataaaattat agattccttt tcagaaaata ttataaagag
39841 gtagataaaa atgaacgagc aaataatagg aagcatatat actttagcag gagggtgtgt
39901 gctttattca gttaaagaga tttttaggta tttacagat ttaaaaaagg ctaaaaaagat
39961 aatcaattta gaacaaatat atccgatata tttagattgt ttaaaaaagg ctaaaaaagat
40021 gattggagct tatattatcc caacagaaca gcatgaattt ttagattttt ttgatattga
40081 agtctttaat aatttagata agcaaatgaa aaaagcgtat gaaaatgtta ttggatttag
40141 acaaatgatt aatttatcaa atagagttaa ggcaatggaa gattttaaga tgagtttcaa
40201 caatgaattt agtacaaatc agattttttt taatccttct tttgttatgg aaacaattgc
40261 tattataaat gaatatcaaa aagatatatc ttatttaaaa aatataatta ataaaaagaa

```

151

```
40321 tgaaaataga gcttataatc atattgatag ttttatcact tcagagtacc gacgaaaaat
40381 aaacgattat aatctttatc ttgataaatt tgaagaacag tttagtcaaa agttttaaatt
40441 aaacagaact tcgataaaaag aaagaattat tattaattta aacaagagga gattttaaatt
40501 atgtggatta ctatgactat tgtatttgct atattgctat tagtttgat cagtattaat
40561 agtgatcgtg caagagagat acaagcactt agatatatga atgattatct acttgatgaa
40621 gtagttaaaa ctaaaagggt caacgggtta gaagaatata ggattgaatt gaagcgaatt
40681 aataacgata ttaaaaagta atttatatta tcggagggtat tgcattgaat gataaagatt
40741 gagaaacacg atatcaaaaa gcttgaagaa tacattcagc acatcgataa ctatcgaaga
40801 gagttgaaga tgcgagaata tgaattactt gaaagtcatg aaccagataa tgcggggagct
40861 ggcaaaagta atttgccggg taacccgatt gaacgatgtg caataaagaa gtttagtgat
40921 aacaggatca atacattaag aaatatagtt aacgggtgtag atagattgat aggtgaaagt
40981 gatgaggata cgcttgagtt attaagggtt agatatggg attgtcctat tgggtgttat
41041 gaattgggaag atatagcaca ttactttggt acaagtaaga caagtatat acgtagaagg
41101 aatgcactga tcgataagtt agcaaaagtat attggttatg ttagcggac tttacccta
41161 tgtaagtccg cattaaaaa gtttattatg ttagtatcag attaatattt aaagtatta
41221 aatgctaata cgacgcatga acaagaggcg catcactatg tgatgtgtct tttatttat
41281 gaggtatgaa catgttcaaa ctaattgtaa atacattact acacatcaag tatagatgag
41341 tcttgatact acttaagtta tataagggtga aacattatga tgactaaaga cgaacgtata
41401 cgattctata agtctaaaga atggcaata acaagaaaaa gagtgtctaga aagagataat
41461 tatgaatgtc aacaatgtaa gagagacggc aagttaacga catatgacaa aagcaagcgt
41521 aagtcgttgg atgtagatca tatattatcg ctagaacatc atccggagtt tgctcatgac
41581 ttaacaatt tagaaacact gtgtattaaa tgtcacaca aaaaagaaaa gagatttata
41641 aaaaagaaa ataaatggaa agacgaaaaa tggtaaatat ccccggtca aaaaatcaa
41701 aagcgatc
```


Table 3

	Name	Position		Name	Position
1	77ORF005	19572..21026	48	77ORF052	1762..2013
2	77ORF006	3976..5196	49	77ORF053	37521..37757
3	77ORF007	21871..23076	50	77ORF054	22818..23060
4	77ORF008	2120..3307	51	77ORF055	17546..17788
5	77ORF009	31946..32803	52	77ORF058	18892..19122
6	77ORF010	26092..26889	53	77ORF059	34564..34785
7	77ORF011	24441..25208	54	77ORF064	29574..29795
8	77ORF012	29788..30576	55	77ORF065	28528..28746
9	77ORF013	33620..34399	56	77ORF066	27494..27703
10	77ORF014	27760..28512	57	77ORF069	38341..38547
11	77ORF015	3291..4028	58	77ORF070	36269..36475
12	77ORF016	32867..33610	59	77ORF071	40498..40701
13	77ORF017	23269..23982	60	77ORF072	38735..38938
14	77ORF018	31169..31840	61	77ORF073	30945..31148
15	77ORF019	39851..40501	62	77ORF074	38544..38738
16	77ORF020	6926..7570	63	77ORF075	13673..13870
17	77ORF021	37762..38304	64	77ORF077	25357..25605
18	77ORF022	30605..31156	65	77ORF079	29089..29280
19	77ORF023	26903..27346	66	77ORF080	35204..35389
20	77ORF024	10700..11140	67	77ORF085	24060..24242
21	77ORF025	9707..10147	68	77ORF092	39706..39876
22	77ORF026	40729..41145	69	77ORF094	32226..32393
23	77ORF027	6518..6925	70	77ORF096	13606..13773
24	77ORF028	34795..35199	71	77ORF098	7092..7256
25	77ORF029	6117..6521	72	77ORF102	29051..29212
26	77ORF030	36478..36879	73	77ORF104	34393..34551
27	77ORF031	39151..39546	74	77ORF109	18282..18434
28	77ORF032	33892..34266	75	77ORF112	39543..39692
29	77ORF033	5758..6120	76	77ORF117	27361..27501
30	77ORF034	7886..8236	77	77ORF118	38390..38530
31	77ORF035	19258..19560	78	77ORF120	36059..36199
32	77ORF036	36876..37223	79	77ORF124	33699..33833
33	77ORF037	102..446	80	77ORF128	14221..14355
34	77ORF038	34908..35219	81	77ORF130	15675..15806
35	77ORF039	37220..37528	82	77ORF133	8414..8542
36	77ORF040	41377..41676	83	77ORF140	13113..13235
37	77ORF041	35454..35753	84	77ORF147	7029..7148
38	77ORF042	5490..5774	85	77ORF149	30668..30787
39	77ORF043	29304..29564	86	77ORF151	31837..31953
40	77ORF044	18481..18768	87	77ORF155	30278..30391
41	77ORF045	5216..5500	88	77ORF157	4044..4157
42	77ORF046	25663..25935	89	77ORF167	20692..20799
43	77ORF047	11159..11425	90	77ORF175	35717..35821
44	77ORF048	28776..29039	91	77ORF176	6836..6940
45	77ORF049	36013..36255	92	77ORF178	35390..35491
46	77ORF050	35753..36007	93	77ORF179	8318..8419
47	77ORF051	38931..39167	94	77ORF182	29268..29564

Table 4

77ORF017 sequence

```

23982      atgacgcataatatagaaaaacgcattaataaattaaaaaacttct
1      M  T  H  N  I  E  K  R  I  N  K  L  K  T  S
23937      ggaaatccaaaattttaaaaagttagattcagatattcactattta
16     G  N  P  K  F  K  K  L  D  S  D  I  H  Y  L
23892      ctcaagagatttgaagggtgaaaaaaaccataaagggtttttatcca
31     L  K  R  F  E  G  E  K  N  H  K  G  F  Y  P
23847      aagtttaaacaggagaaatagttttttagatttcggtataaac
46     K  F  K  Q  G  E  I  V  F  V  D  F  G  I  N
23802      gttaataaagaatttttctaattcacactttgcaatagtgatgaat
61     V  N  K  E  F  S  N  S  H  F  A  I  V  M  N
23757      aaaaatgattctaatacggaggatatagtaaattgttattccctta
76     K  N  D  S  N  T  E  D  I  V  N  V  I  P  L
23712      tcctctaagaaaaacaaaaagtattttaagatgaattttgatttg
91     S  S  K  E  N  K  K  Y  L  K  M  N  F  D  L
23667      aaatgggagtattatttaagattgttttttaaatttaattagcgcg
106    K  W  E  Y  Y  L  R  L  F  L  N  L  I  S  A
23622      caaaataattcagctatattaaaagaagttttcgataaaaaatac
121    Q  N  N  S  A  I  L  K  E  V  F  D  K  K  Y
23577      caaaaaaacacacagaattcatcactaaagattattttattgaa
136    Q  K  N  N  T  E  F  I  T  K  D  Y  F  I  E
23532      tttatatctgatagtttagaaattgaaaataaattaaataaaatt
151    F  I  S  D  S  L  E  I  E  N  K  L  N  K  I
23487      gacagaaacattaataacatagtatcagcaattgataaggtaaaa
166    D  R  N  I  N  N  I  V  S  A  I  D  K  V  K
23442      aaattaaaaggtaatagttacgcttgcataaattctttccagccg
181    K  L  K  G  N  S  Y  A  C  I  N  S  F  Q  P
23397      attagtaagtttcgcataagaaaagttttacccccaaaaaattaaa
196    I  S  K  F  R  I  R  K  V  L  P  Q  K  I  K
23352      aatccagtaatagattcttcggatattatgttactgataaataga
211    N  P  V  I  D  S  S  D  I  M  L  L  I  N  R
23307      attaataataatatattgcagatccctgatataagatga 23269
226    I  N  N  N  I  L  Q  I  P  D  I  R  *

```

Physico-chemical parameters of ORF 77ORF017

```

1      MTHNIEKRIN KLKTSQNPKE KKLDSDIHYL LKRFEQEKNH KGFYKPKQGG EIVFVDFGIN
61     VNKEFSNSHF AIVMNKNSDN TEDIVNVIPL SSKENKKYLK MNFDLKWEYY LRLFLNLISA
121    QNNSAILKEV FDKKYQKQNT EFITKDYFIE FISDSLEIEN KLNKIDRNIN NIVSAIDKVK
181    KLKGSYACI NSFQISKFR IRKVLQKIK NPVIDSSDIM LLINRINNNI LQIPDIR

```

Number of amino acids: 237
 Average molecular weight (Daltons): 27887.38
 Mean amino acid weight (Daltons): 117.67
 Monoisotopic molecular weight (Daltons): 27869.83
 Mean amino acid monoisotopic weight (Daltons): 117.59

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	5	2.11%	7.58%	Cys	C	1	0.42%	1.66%
Asp	D	14	5.91%	5.28%	Glu	E	13	5.49%	6.37%
Phe	F	16	6.75%	4.09%	Gly	G	6	2.53%	6.84%
His	H	4	1.69%	2.24%	Ile	I	29	12.24%	5.81%
Lys	K	33	13.92%	5.95%	Leu	L	19	8.02%	9.42%
Met	M	4	1.69%	2.37%	Asn	N	30	12.66%	4.45%
Pro	P	7	2.95%	4.9%	Gln	Q	6	2.53%	3.97%
Arg	R	8	3.38%	5.16%	Ser	S	17	7.17%	7.12%
Thr	T	5	2.11%	5.67%	Val	V	11	4.64%	6.58%
Trp	W	1	0.42%	1.23%	Tyr	Y	8	3.38%	3.18%

Number of acidic (negative) amino acids (ED): 27
 11.39%
 Number of basic (positive) amino acids (KR): 41
 17.30%
 Total charge (KRED): 68
 28.69%
 Net charge (KR - ED): 14
 5.91%
 Theoretical pI: 10.01
 Total linear charge density: 0.30
 Average hydrophobicity: -5.37
 Ratio of hydrophilicity to hydrophobicity: 1.41
 Percentage of hydrophilic amino acid: 57.81%
 Percentage of hydrophobic amino acid: 42.19%
 Ratio of %hydrophilic to %hydrophobic: 1.37

77ORF019 sequence

```
39851      atgaacgagcaaataataggaagcatatatacttttagcaggaggt
1   M N E Q I I G S I Y T L A G G
39896      gttgtgctttattcagttaaagagattttttaggtatttttacagat
16  V V L Y S V K E I F R Y F T D
39941      tctaacttacaacgtaaaaaaatcaatttagaacaaatatatccg
31  S N L Q R K K I N L E Q I Y P
39986      atatatatttagattgttttaaaaaggctaaaaagatgattggagct
46  I Y L D C F K K A K K M I G A
40031      tatattattccaacagaacagcatgaatttttagatttttttgat
61  Y I I P T E Q H E F L D F F D
40076      attgaagtctttaataatttagataagcaaagtaaaaaagcgtat
76  I E V F N N L D K Q S K K A Y
40121      gaaaatgttattggatttagacaaatgattaatttatcaaataga
91  E N V I G F R Q M I N L S N R
40166      gttaaggcaatggaagattttaagatgagtttcaacaatgaattt
106 V K A M E D F K M S F N N E F
40211      agtacaaatcagattttttttaatccttcttttgttatggaaaca
121 S T N Q I F F N P S F V M E T
40256      attgctattataaatgaatatcaaaaagatatatcttatttataaa
136 I A I I N E Y Q K D I S Y L K
40301      aatataattaataaaaatgaatgaaaatagagcttataatcatatt
151 N I I N K M N E N R A Y N H I
40346      gatagtttttatcacttcagagtaccgacgaaaaataaacgattat
166 D S F I T S E Y R R K I N D Y
40391      aatcttttatcttgataaatttgaagaacagtttagtcaaaaagttt
181 N L Y L D K F E E Q F S Q K F
40436      aaaataaacagaacttcgataaaagaaagaattattattaattta
196 K I N R T S I K E R I I I N L
40481      aacaagaggagattttaaatga 40501
211 N K R R F K *
```

Physico-chemical parameters of ORF 77ORF019

```

1      MNEQIIGSIY TLAGGVVLYS VKEIFRYFTD SNLQRKKINL EQIYPIYLDK FKKAKKMIGA
61     YIIPTEQHEF LDFFDIEVFN NLDKQSKKAY ENVIGFRQMI NLSNRVKAME DFKMSFNNEF
121    STNQIFFNPS FVMETIAIIN EYQKDISYK NIINKMNENR AYNHIDSFIT SEYRRKINDY
181    NLYLDKFEEQ FSQKFKINRT SIKERIIINL NKRRFK

```

Number of amino acids: 216
 Average molecular weight (Daltons): 26026.06
 Mean amino acid weight (Daltons): 120.49
 Monoisotopic molecular weight (Daltons): 26009.34
 Mean amino acid monoisotopic weight (Daltons): 120.41

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	7	3.24%	7.58%	Cys	C	1	0.46%	1.66%
Asp	D	10	4.63%	5.28%	Glu	E	16	7.41%	6.37%
Phe	F	19	8.80%	4.09%	Gly	G	5	2.31%	6.84%
His	H	2	0.93%	2.24%	Ile	I	28	12.96%	5.81%
Lys	K	22	10.19%	5.95%	Leu	L	12	5.56%	9.42%
Met	M	7	3.24%	2.37%	Asn	N	23	10.65%	4.45%
Pro	P	3	1.39%	4.9%	Gln	Q	10	4.63%	3.97%
Arg	R	11	5.09%	5.16%	Ser	S	13	6.02%	7.12%
Thr	T	7	3.24%	5.67%	Val	V	7	3.24%	6.58%
Trp	W	0	0.00%	1.23%	Tyr	Y	13	6.02%	3.18%

Number of acidic (negative) amino acids (ED): 26
 12.04%
 Number of basic (positive) amino acids (KR): 33
 15.28%
 Total charge (KRED): 59
 27.31%
 Net charge (KR - ED): 7
 3.24%
 Theoretical pI: 9.52
 Total linear charge density: 0.28
 Average hydrophobicity: -4.84
 Ratio of hydrophilicity to hydrophobicity: 1.37
 Percentage of hydrophilic amino acid: 54.17%
 Percentage of hydrophobic amino acid: 45.83%
 Ratio of %hydrophilic to %hydrophobic: 1.18

77ORF043 sequence

```
29304      atgtattacgaaataggcgaaatcatacgcaaaaatattcatgtt
1      M Y Y E I G E I I R K N I H V
29349      aacggattcgattttaagctattcatttttaaagggtcatatgggc
16     N G F D F K L F I L K G H M G
29394      atatcaatacaagttaaagatatgaacaacgtaccaattaaacat
31     I S I Q V K D M N N V P I K H
29439      gcttatgtcgtagatgagaatgacttagatatggcatcagactta
46     A Y V V D E N D L D M A S D L
29484      ttttaaccaagcaatagatgaatggattgaagagaacacagacgaa
61     F N Q A I D E W I E E N T D E
29529      caggacagactaattaacttagtcatgaaatggtag 29564
76     Q D R L I N L V M K W *
```

Physico-chemical parameters of ORF 77ORF043

1 MYYEIGEIIIR KNIHVNGFDF KLFILKGHMG ISIQVKDMNN VPIKHAYVVD ENLDLMASDL
61 FNQAIDEWIE ENTDEQDRLI NLVMKW

Number of amino acids: 86
Average molecular weight (Daltons): 10186.68
Mean amino acid weight (Daltons): 118.45
Monoisotopic molecular weight (Daltons): 10180.02
Mean amino acid monoisotopic weight (Daltons): 118.37

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	3	3.49%	7.58%	Cys	C	0	0.00%	1.66%
Asp	D	9	10.47%	5.28%	Glu	E	7	8.14%	6.37%
Phe	F	4	4.65%	4.09%	Gly	G	4	4.65%	6.84%
His	H	3	3.49%	2.24%	Ile	I	11	12.79%	5.81%
Lys	K	6	6.98%	5.95%	Leu	L	6	6.98%	9.42%
Met	M	5	5.81%	2.37%	Asn	N	8	9.30%	4.45%
Pro	P	1	1.16%	4.9%	Gln	Q	3	3.49%	3.97%
Arg	R	2	2.33%	5.16%	Ser	S	2	2.33%	7.12%
Thr	T	1	1.16%	5.67%	Val	V	6	6.98%	6.58%
Trp	W	2	2.33%	1.23%	Tyr	Y	3	3.49%	3.18%

Number of acidic (negative) amino acids (ED): 16
18.60%
Number of basic (positive) amino acids (KR): 8
9.30%
Total charge (KRED): 24
27.91%
Net charge (KR - ED): -8
9.30%
Theoretical pI: 4.38
Total linear charge density: 0.30
Average hydrophobicity: -2.80
Ratio of hydrophilicity to hydrophobicity: 1.19
Percentage of hydrophilic amino acid: 48.84%
Percentage of hydrophobic amino acid: 51.16%
Ratio of %hydrophilic to %hydrophobic: 0.95

77ORF102 sequence

```
29051      atgagcaacattttataaaaagctacctagtagcagtattatgcttc
1      M  S  N  I  Y  K  S  Y  L  V  A  V  L  C  F
29096      acagtcttagcgattgtacttatgccgtttctatacttcactaca
16     T  V  L  A  I  V  L  M  P  F  L  Y  F  T  T
29141      gcatgggtcaattgcgggattcgcaagtatcgcaacattcatgtac
31     A  W  S  I  A  G  F  A  S  I  A  T  F  M  Y
29186      taaaagaatgctttttcaagaataa 29212
46     Y  K  E  C  F  F  K  E  *
```

Physico-chemical parameters of ORF 77ORF102

1 MSNIYKSYLV AVLCTVLAI VLMPLYFTT AWSIAGFASI ATFMYYKECF FKE

Number of amino acids: 53
 Average molecular weight (Daltons): 6155.42
 Mean amino acid weight (Daltons): 116.14
 Monoisotopic molecular weight (Daltons): 6151.07
 Mean amino acid monoisotopic weight (Daltons): 116.06

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	6	11.32 %	7.58%	Cys	C	2	3.77 %	1.66%
Asp	D	0	0.00%	5.28%	Glu	E	2	3.77 %	6.37%
Phe	F	7	13.21 %	4.09%	Gly	G	1	1.89 %	6.84%
His	H	0	0.00%	2.24%	Ile	I	4	7.55 %	5.81%
Lys	K	3	5.66%	5.95%	Leu	L	5	9.43 %	9.42%
Met	M	3	5.66%	2.37%	Asn	N	1	1.89 %	4.45%
Pro	P	1	1.89%	4.9%	Gln	Q	0	0.00 %	3.97%
Arg	R	0	0.00%	5.16%	Ser	S	4	7.55 %	7.12%
Thr	T	4	7.55%	5.67%	Val	V	4	7.55 %	6.58%
Trp	W	1	1.89%	1.23%	Tyr	Y	5	9.43 %	3.18%

Number of acidic (negative) amino acids (ED): 2
 3.77%
 Number of basic (positive) amino acids (KR): 3
 5.66%
 Total charge (KRED): 5
 9.43%
 Net charge (KR - ED): 1
 1.89%
 Theoretical pI: 8.18
 Total linear charge density: 0.13
 Average hydrophobicity: 10.81
 Ratio of hydrophilicity to hydrophobicity: 0.40
 Percentage of hydrophilic amino acid: 28.30%
 Percentage of hydrophobic amino acid: 71.70%

Ratio of %hydrophilic to %hydrophobic:

0.39

77ORF104 sequence

```
34393      atggtaaccaaagaatttttaaaaactaaacttgagtggttcagat
1   M   V   T   K   E   F   L   K   T   K   L   E   C   S   D
34438      atgtacgctcagaaaactcatagatgaggcacagggcgatgaaaat
16  M   Y   A   Q   K   L   I   D   E   A   Q   G   D   E   N
34483      aggttgtagcaccctatttatccaaaaacttgcagaacgtcataca
31  R   L   Y   D   L   F   I   Q   K   L   A   E   R   H   T
34528      cgccccgctatcgtcgaatattaa 34551
46  R   P   A   I   V   E   Y   *
```

Physico-chemical parameters of ORF 77ORF104

1 MVTKEFLKTK LECSDMYAQK LIDEAQGDEN RLYDLFIQKL AERHTRPAIV EY

Number of amino acids: 52
 Average molecular weight (Daltons): 6193.13
 Mean amino acid weight (Daltons): 119.10
 Monoisotopic molecular weight (Daltons): 6189.12
 Mean amino acid monoisotopic weight (Daltons): 119.02

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	4	7.69 %	7.58%	Cys	C	1	1.92%	1.66%
Asp	D	4	7.69 %	5.28%	Glu	E	6	11.54 %	6.37%
Phe	F	2	3.85 %	4.09%	Gly	G	1	1.92%	6.84%
His	H	1	1.92 %	2.24%	Ile	I	3	5.77%	5.81%
Lys	K	5	9.62 %	5.95%	Leu	L	6	11.54 %	9.42%
Met	M	2	3.85 %	2.37%	Asn	N	1	1.92%	4.45%
Pro	P	1	1.92 %	4.9%	Gln	Q	3	5.77%	3.97%
Arg	R	3	5.77 %	5.16%	Ser	S	1	1.92%	7.12%
Thr	T	3	5.77 %	5.67%	Val	V	2	3.85%	6.58%
Trp	W	0	0.00 %	1.23%	Tyr	Y	3	5.77%	3.18%

Number of acidic (negative) amino acids (ED): 10
 19.23%
 Number of basic (positive) amino acids (KR): 8
 15.38%
 Total charge (KRED): 18
 34.62%
 Net charge (KR - ED): -2
 3.85%
 Theoretical pI: 5.03
 Total linear charge density: 0.38
 Average hydrophobicity: -5.81
 Ratio of hydrophilicity to hydrophobicity: 1.47
 Percentage of hydrophilic amino acid: 53.85%
 Percentage of hydrophobic amino acid: 46.15%

Ratio of %hydrophilic to %hydrophobic:

1.17

77ORF182 sequence

```
29268      atgttcaatataaaaacgaaaaacggaggaagtcaagatgtattac
1      M F N I K R K T E E V K M Y Y
29313      gaaataggcgaaatcatacgcaaaaatattcatgttaacggattc
16     E I G E I I R K N I H V N G F
29358      gattttaagctattcattttaaaagggtcatatgggcatatcaata
31     D F K L F I L K G H M G I S I
29403      caagttaaagatatgaacaacgtaccaattaacatgcttatgtc
46     Q V K D M N N V P I K H A Y V
29448      gtagatgagaatgacttagatatggcatcagacttatttaaccaa
61     V D E N D L D M A S D L F N Q
29493      gcaatagatgaatggattgaagagaacacagacgaacaggacaga
76     A I D E W I E E N T D E Q D R
29538      ctaattaacttagtcatgaaatggtag 29564
91     L I N L V M K W *
```

Physico-chemical parameters of ORF 77ORF182

1 MFNIKRKTEE VKMYEIGEI IRKNIHVNGF DFKLFILKGH MGISIQVKDM NNVPIKHAYV
 61 VDENDLDMAS DLFNQAIDEW IEENTDEQDR LINLVMKW

Number of amino acids: 98
 Average molecular weight (Daltons): 11691.50
 Mean amino acid weight (Daltons): 119.30
 Monoisotopic molecular weight (Daltons): 11683.84
 Mean amino acid monoisotopic weight (Daltons): 119.22

Amino acid composition

Aci d	Symbo l	Numb er	%	Average % in Swissprot	Aci d	Symbo l	Numb er	%	Average % in Swissprot
Ala	A	3	3.06 %	7.58%	Cys	C	0	0.00%	1.66%
Asp	D	9	9.18 %	5.28%	Glu	E	9	9.18%	6.37%
Phe	F	5	5.10 %	4.09%	Gly	G	4	4.08%	6.84%
His	H	3	3.06 %	2.24%	Ile	I	12	12.24 %	5.81%
Lys	K	9	9.18 %	5.95%	Leu	L	6	6.12%	9.42%
Met	M	6	6.12 %	2.37%	Asn	N	9	9.18%	4.45%
Pro	P	1	1.02 %	4.9%	Gln	Q	3	3.06%	3.97%
Arg	R	3	3.06 %	5.16%	Ser	S	2	2.04%	7.12%
Thr	T	2	2.04 %	5.67%	Val	V	7	7.14%	6.58%
Trp	W	2	2.04 %	1.23%	Tyr	Y	3	3.06%	3.18%

Number of acidic (negative) amino acids (ED): 18
 18.37%
 Number of basic (positive) amino acids (KR): 12
 12.24%
 Total charge (KRED): 30
 30.61%
 Net charge (KR - ED): -6
 6.12%
 Theoretical pI: 4.76
 Total linear charge density: 0.33
 Average hydrophobicity: -3.89
 Ratio of hydrophilicity to hydrophobicity: 1.28

Percentage of hydrophilic amino acid:	51.02%
Percentage of hydrophobic amino acid:	48.98%
Ratio of %hydrophilic to %hydrophobic:	1.04

Table 5

BLASTP 2.0.8 (Jan-05-1999)

Query= sid|100017|lan|77ORF017 Phage 77 ORF |23269-23982|-3
(237 letters)

Database: nr
393,678 sequences; 120,452,765 total letters

Sequences producing significant alignments:		Score (bits)	E Value
gi 4493986 emb CAB39045.1	(AL034559) predicted using hexExon; ...	41	0.010
gi 730607 sp P23250 RPI1_YEAST	NEGATIVE RAS PROTEIN REGULATOR P...	38	0.053
gi 3097044 emb CAA75299	(Y15035) K1R [Cowpox virus]	38	0.090
gi 2146245 pir S73794	hypothetical protein H91_orf180 - Mycopl...	38	0.090
gi 83910 pir S04682	ribosomal protein var1 - yeast (Candida gl...	37	0.15
gi 133135 sp P21358 RMAR_CANGA	MITOCHONDRIAL RIBOSOMAL PROTEIN ...	37	0.15
gi 2128843 pir H64475	hypothetical protein MJ1409 - Methanococ...	36	0.20
gi 5107017 gb AAD39926.1 AF126285_2	(AF126285) RNA polymerase {...	36	0.35
gi 2146210 pir S73342	hypothetical protein E07_orf166 - Mycopl...	35	0.60

Database: swissprot
79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:		Score (bits)	E Value
sp P23250 RPI1_YEAST	NEGATIVE RAS PROTEIN REGULATOR PROTEIN.	38	0.014
sp P21358 RMAR_CANGA	MITOCHONDRIAL RIBOSOMAL PROTEIN VAR1.	37	0.040
sp Q21444 LDLC_CAEEL	LDLC PROTEIN HOMOLOG.	34	0.35
sp P27240 RFAY_ECOLI	LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS PROT.	33	0.46
sp P53192 YGCO_YEAST	HYPOTHETICAL 27.1 KD PROTEIN IN ALX1-CXB1.	33	0.60
sp P32908 SMC1_YEAST	CHROMOSOME SEGREGATION PROTEIN SMC1 (DA-B.	33	0.60
sp P54683 TAGB_DICDI	PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR .	32	0.78
sp Q03100 CYAA_DICDI	ADENYLATE CYCLASE, AGGREGATION SPECIFIC (.	32	0.78

169

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|100019|lan|77ORF019 Phage 77 ORF|39851-40501|2
(216 letters)

Database: nr
373,355 sequences; 114,214,446 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 3341966 dbj BAA31932 (AB009866) orf 59 [bacteriophage phi PVL]	437	e-122
gi 2689911 (AE000792) B. burgdorferi predicted coding region BB...	38	0.058
gi 1171589 emb CAA64574 (X95275) frameshift [Plasmodium falcip...	37	0.10
gi 4493986 emb CAB39045.1 (AL034559) predicted using hexExon; ...	36	0.23
gi 141257 sp P18019 YPI9_CLOPE HYPOTHETICAL 14.5 KD PROTEIN (OR...	36	0.29
gi 133412 sp P27059 RPOB_ASTLO DNA-DIRECTED RNA POLYMERASE BETA...	35	0.51
gi 3122231 sp Q58851 HISX_METJA HISTIDINOL DEHYDROGENASE (HDH) ...	35	0.51
gi 3649757 emb CAB11106.1 (Z98547) predicted using hexExon; MA...	34	0.66
gi 2688313 (AE001146) sensory transduction histidine kinase, pu...	34	0.87

Database: swissprot
79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P18019 YPI9_CLOPE HYPOTHETICAL 14.5 KD PROTEIN (ORF9).	36	0.079
sp Q58851 HISX_METJA HISTIDINOL DEHYDROGENASE (EC 1.1.1.23) (H.	35	0.14
sp P27059 RPOB_ASTLO DNA-DIRECTED RNA POLYMERASE BETA CHAIN (E.	35	0.14
sp Q02224 CENE_HUMAN CENTROMERIC PROTEIN E (CENP-E PROTEIN).	34	0.31
sp P04931 ARP_PLAFA ASPARAGINE-RICH PROTEIN (AG319) (ARP) (FRA..	33	0.53
sp P18011 IPAB_SHIFL 62 KD MEMBRANE ANTIGEN.	32	0.69
sp P18709 VTA2_XENLA VITELLOGENIN A2 PRECURSOR (VTG A2) (CONTA..	32	0.90
sp Q64409 CP3H_CAVPO CYTOCHROME P450 3A17 (EC 1.14.14.1) (CYPI..	32	0.90
sp P21358 RMAR_CANGA MITOCHONDRIAL RIBOSOMAL PROTEIN VAR1.	32	0.90
sp Q03945 IPAB_SHIDY 62 KD MEMBRANE ANTIGEN.	32	1.2

170

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|100043|lan|77ORF043 Phage 77 ORF|29304-29564|3
(86 letters)

Database: nr
373,355 sequences; 114,214,446 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 3341947 dbj BAA31913 (AB009866) orf 39 (bacteriophage phi PVL)	182	6e-46
gi 744518 prf 2014422A FKBP-rapamycin-associated protein [Homo...	32	0.84
gi 1169736 sp P42346 FRAP_RAT FKBP-RAPAMYCIN ASSOCIATED PROTEIN...	32	0.84
gi 1169735 sp P42345 FRAP_HUMAN FKBP-RAPAMYCIN ASSOCIATED PROTE...	32	0.84
gi 3282239 (U88966) rapamycin associated protein FRAP2 (Homo sa...	32	0.84
gi 3875402 emb CAA98122 (Z73906) cDNA EST EMBL:D64544 comes fr...	31	2.5
gi 1084792 pir S54091 hypothetical protein YPR070w - yeast (Sa...	30	4.2

Database: swissprot
79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P42345 FRAP_HUMAN FKBP-RAPAMYCIN ASSOCIATED PROTEIN (FRAP) .	32	0.24
sp P42346 FRAP_RAT FKBP-RAPAMYCIN ASSOCIATED PROTEIN (FRAP) (R.	32	0.24
sp P34554 YNP1_CAEEL HYPOTHETICAL 42.2 KD PROTEIN T05G5.1 IN C.	28	3.5
sp Q24118 LIO_DROME LINOTTE PROTEIN.	28	3.5
sp P80034 ACH2_BOMMO ANTICHYMOTRYPSIN II (ACHY-II).	28	3.5
sp P22922 ALAT_BOMMO ANTITRYPSIN PRECURSOR (AT).	28	3.5
sp Q44363 TRAA_AGR6 CONJUGAL TRANSFER PROTEIN TRAA.	28	3.5
sp P38255 YBUS_YEAST HYPOTHETICAL 51.3 KD PROTEIN IN PHO5-VPS1.	27	6.0
sp P55822 SH3B_HUMAN SH3BGR PROTEIN (21-GLUTAMIC ACID-RICH PRO.	27	7.9
sp Q58482 YA82_METJA HYPOTHETICAL PROTEIN MJ1082.	27	7.9
sp P34252 YKK8_YEAST HYPOTHETICAL 52.3 KD PROTEIN IN HAP4-AAT1.	27	7.9

171

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|100102|lan|77ORF102 Phage 77 ORF|29051-29212|2
(53 letters)

Database: nr
373,355 sequences; 114,214,446 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 3341946 dbj BAA31912 (AB009866) orf 38 [bacteriophage phi PVL]	96	3e-20
gi 4325288 gb AAD17315 (AF123593) voltage-dependent sodium cha...	28	7.1
gi 2649684 (AE001040) A. fulgidus predicted coding region AF092...	28	9.3

Database: swissprot
79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P42087 HUTM_BACSU PUTATIVE HISTIDINE PERMEASE.	26	7.1
sp P04775 CIN2_RAT SODIUM CHANNEL PROTEIN, BRAIN II ALPHA SUBU...	26	9.2
sp P42619 YQJF_ECOLI HYPOTHETICAL 17.2 KD PROTEIN IN EXUR-TDCC...	26	9.2

172

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|100104|lan|77ORF104 Phage 77 ORF|34393-34551|1
(52 letters)

Database: nr
373,355 sequences; 114,214,446 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 2315523 (AF016452) similar to the leucine-rich domains found...	29	4.2
gi 4377168 gb AAD18990 (AE001666) CT711 hypothetical protein [...	29	5.4
gi 3882171 dbj BAA34445 (AB018268) KIAA0725 protein [Homo sapi...	28	9.3

Database: swissprot
79,449 sequences; 28,874,452 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P04879 RRPP_VSVIG RNA POLYMERASE ALPHA SUBUNIT (EC 2.7.7.48.	27	5.4
sp P04880 RRPP_VSVIM RNA POLYMERASE ALPHA SUBUNIT (EC 2.7.7.48.	27	5.4
sp Q13946 CN7A_HUMAN HIGH-AFFINITY CAMP-SPECIFIC 3',5'-CYCLIC .	26	7.1
sp P35381 ATPA_DROME ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL P.	26	9.3
sp P54659 MVPB_DICDI MAJOR VAULT PROTEIN BETA (MVP-BETA).	26	9.3
sp P40397 YHXC_BACSU HYPOTHETICAL OXIDOREDUCTASE IN APRE-COMK .	26	9.3

173

BLASTP 2.0.8 [Jan-05-1999]

Query= sid|122748|lan|77ORF182 Phage 77 ORF|29268-29564|3
(98 letters)

Database: nr
393,678 sequences; 120,452,765 total letters

Sequences producing significant alignments:	Score (bits)	E Value
gi 3341947 dbj BAA31913.1 (AB009866) orf 39 [bacteriophage phi..	182	8e-46
gi 1084792 pir S54091 hypothetical protein YPR070w - yeast (Sa..	35	0.13
gi 1169736 sp P42346 FRAP_RAT FKBP-RAPAMYCIN ASSOCIATED PROTEIN..	32	1.1
gi 744518 prf 2014422A FKBP-rapamycin-associated protein [Homo..	32	1.1
gi 5051381 emb CAB44736.1 (AL049653) dJ647M16.2 (FK506 binding..	32	1.1
gi 4826730 ref NP_004949.1 pFRAP1 FK506 binding protein 12-rap..	32	1.1
gi 3282239 (U88966) rapamycin associated protein FRAP2 [Homo sa..	32	1.1

Database: swissprot
79,909 sequences; 29,054,478 total letters

Sequences producing significant alignments:	Score (bits)	E Value
sp P42345 FRAP_HUMAN FKBP-RAPAMYCIN ASSOCIATED PROTEIN (FRAP) .	32	0.29
sp P42346 FRAP_RAT FKBP-RAPAMYCIN ASSOCIATED PROTEIN (FRAP) (R.	32	0.29
sp P40557 YIA5_YEAST PUTATIVE DISULFIDE ISOMERASE YIL005W PREC.	29	3.3
sp Q24118 LIO_DROME LINOTTE PROTEIN.	28	4.4
sp Q44363 TRAA_AGRT6 CONJUGAL TRANSFER PROTEIN TRAA.	28	4.4
sp P80034 ACH2_BOMMO ANTICHYMOTRYPSIN II (ACHY-II).	28	4.4
sp P34554 YNP1_CAEEL HYPOTHETICAL 42.2 KD PROTEIN T05G5.1 IN C.	28	4.4
sp P22922 A1AT_BOMMO ANTITRYPSIN PRECURSOR (AT).	28	4.4

Table 6

1st position (5' end)	2nd position				3rd position (3' end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Table 7

Bacteriophage 3A, complete genome sequence

```

1      caaacgctag caacgcggat aaatttttca tgaagggggg tcttttatatg aagttaacaa aaaaacagct
71     aaaaagatat atagaagatt acaaaaaatc tgaatgacata ttaatttaatt tgtatataga aacaratgaa
141    ttttattgtc ggttaagaga tgaacttaaa aatagtgtat taatgataga gcatacaaac aaggctgggtg
211    cgagcaatat tattaagaat ccattaagca tagaactgac aaaaacagtt caaacactaa ataacttact
281    caagtctatg ggtttaactg cagcacaaag aaaaagata gttcaagaag aaggtggatt cgggtgactat
351    taaagtttta aatgaacctt caccaaaact attaacaca tggtatgcag agcaagtcac tcaagggaaa
421    ataaaaacaa gcaaatatgt tagaaaagaa tgcgagagac atcttagata tctagaaaat ggaggttaaat
491    ggggtatttga tgaagaatta ggcgcatcgt cttatcgatt tatagaaaag ttttgttaac cttccaaagg
561    atctaaacgt caacttgtat tacagccatg gcaacatttt attatcggca gtttgccttg tgggttcct
631    aagaaacaaa aactgcgcag gtttaaaaga gctttgatat ttatggggcg aaaaaatcgt aaaaacacca
701    ctattttctg ggttgccta ctagctgtat cacaagatgg agaaaatcgt gcagaattc atttgttagc
771    aaacgtaagt aaacaagcta ggatttctatt tgaatgaatt aagggcagta ttaaaagctag cccaaagctt
841    gataaaaatt tcagaacatt aagagatgaa atccattatg acgcaacgat atcaaaaatt atgccccaa
911    catcagatat cgataagtta gatggattga atacacacat ggggattttt gatgaattc atgaatttaa
981    agactataaa ttgatttcag ttataaaaaa ctcaagagct gcaagggtac aacctcttct catctacatt
1051   acgacagcag ggtatcaatt agatgggtcca cttgttgata tggtagaagc ggggaagagc accttagatc
1121   aaatcattaga agcgaagaa accttttact atttagcatt ttggatgat gacgatgata ttaagtattc
1191   gtcgaactgg ataaaaacaa atcccaactt aggtgtctct ataaatttag atgagatgaa agaagagtg
1261   gaaaaagcta agagaacacc agctgaacgt ggagatttta taacaaaag gtttaatatc tttgctaata
1331   atgacgagat gagttttatt gattacccaa cactccaaaa aaataatgaa attgtttctt tagaagagct
1401   ggaaggcaga ccgtgcacga ttggttatga tttatcagaa acagaggact ttacagccgc gttggtcact
1471   tttgcgctag ataatggtaa agttgcagtt ttatcgcatc catggattcc taagcacaaa gttgaattt
1541   ctaacgaaaa aataccctat agagaatggg aagaagatgg cttattaaca gtcgaagata agccttatat
1611   tgactaccaa gatgttttaa attggtataa taagatgaat gacgattatg tagtagaaaa aattacttat
1681   gatagagcga cgcattcaa actaaatcaa gagttaaaaa attacgggtt tgaacgggaa gaaacaagac
1751   aaggagcttt gaccttgagc cctgcattga aggatttaaa agaaatggtt ttagatggga aaataattt
1821   taataataat cctttaatga aatgggtatat caataatgtt cagttgaaac tagacagaaa cggaaactgg
1891   ttgcccgtta agcaaacgag atatcgtaaa atagatggct ttgcagcatt tttaaacaca tatacagata
1961   ttatgaataa agttgtttct gatagtgggt aaggaaacat agagtttatt agtattaaag acataactgc
2031   ttaaggaggt gaattgtatc gcaaaagaga atattgtcac acgcataaag aaaaaattga tagacaattg
2101   gattgattcag tcaacttcta agctttatga ctttagccca tggaaaaata gatcttttgg ggggtgtaatt
2171   aataaaccgc ttgaaactaa tgaacagata ttttcagcta ttacaaagtt atctaattcg atggctagtt
2241   tgcctctgaa atcgtatgaa gattataaag tagttaaatc agaagtatct gatttactta cagtgtcacc
2311   gaataattct ctgagcaggt ttgattttat taatcaaat gaaacaatca gaaatgaaaa aggtaatgca
2381   tatgtgtcaa ttgaacgaga catctatcat caaccatcaa agcttttctt attaaatcca gatgtgttg
2451   aaatgttaat tgaaaaccaa tcacgtgaac tttattattc catctatgct gcaactggaa ataaattgat
2521   tgttcataat atggacatgt tgcattttaa acacatcgtg gcatcctaata tgggtgcaagg cattagtccg
2591   attgatgtgt tgaagaatac aactgatttt gataatgcag taagaacctt taatcttaca gaaatgcaaa
2661   aacctgattc tttcatgctt aaatatgggt ccaatgtagg taagaaaaaa aggcagcaag tgttagaaga
2731   tttcaaacag tactatgaag aaaaacggtg aatattattc caagagcctg gttgtgaaat cgaacggtta
2801   cctaaaaaat atgtctctga agatatagtg gcaagcgaga atttaacaag agaaagagta gctaactgtt
2871   ttcaattgct ctcagtratt ttaaatgcaa gatcaaatac aaatttcgct aaaaatgaag agttaaacag
2941   atttactctg cagcatacct tattgccaat cgtcaaacag tatgaagaag aatttaactg gaaactactt
3011   actaaaaacg acagagaaaa aaatagggtat ttttaattta acgtttaaatt ctatttaagg gctgatagtg
3081   caacacaagc agaagtgatc ttaaaagcag ttctgtagtg ttactacact ataaatgaca ttagagagtg
3151   ggaagattta ccaccagrtg aaggtggaga taagccgcta ataagcggtg atttataccc aattgacacg
3221   ccacttgaat taagaaaaatc tttagaaagg ggtgataaaa atgtcaatga aagctaagta ttttcaaatg
3291   aaaaagaaat caaaaagtaa aggtgaaata tttatttatg gtgatattgt aagtgataaa tgggttgaaa
3361   gtgagttaac tgcacagat ttcaaaaaa aactagatga actaggagac atcagtgaat tagatgttca
3431   tataaattca tctggaggca gtgtatttga agggcatgca atatacaata tgcataaaat gcattctgca
3501   aaaaattata tctatgtcga tgccttagcg gcatcaattg ctagtgttat cgtcatgagt ggtgacacta
3571   tttttatgca caaaaatagt tttttaatga ttcataattc atgggttatg actgtaggta atgcagaaga
3641   gtttaagaaag acagcggatt tacttgaaaa aacagatgct gttagttaatt cagcttattt agataaagca
3711   aagattttag atcaagaaca cttaaaacag atgttagatg cagaactttg gcttactgca gaagaagcct
3781   tgtctttcgg cttgtagatg gaaatttttag gagctaattga aataactgct agtatctcta aagagcaata
3851   taagcgtttc gagaacgtcc cagaagattt aaagaaagat gttagcaaaa tcaactaaat ccatgtatga
3921   gatcgtttg aattggttga aacacctaaa gaaagtatgt cactagaaga aaaaagaaaa agagaaaaaa
3991   ttaaacgcga atgcgaattt ttaaaaatga caatgagtta ttaggaggaa atgaaaatgcc gacattatat
4061   gaattaaac aatccttagg tatgattgga caacaatraa aaaaataaaa tgatgaattg agtcagaag
4131   caaccagacc aaaaattgat atggaagaca tcaaacact agaaacagaa aaagcaggct tacaacaaag
4201   atttaacatt gttgaaagac aagtaaaaga cattgaagaa aaagaaaaag cgaaggttaa agcacagga
4271   gaagcttacc aatctttaa tgatcatgag aagatgggtt aagctaaggc agagttttat cgtcacgga
4341   ttttaccaaa tgaatttgaa aaaccttcaa tggaggcaca acgtttatta cacgctttac caacaggtta
4411   tgattcaggt ggtgataagc tcttaccaaa aacactttct aaagaaattg tttcagaacc atttgttaaa
4481   aaccaattac gtgaaaaagc tctgctaact aacactaaag gtttagagat tccaagagtt tcatataatt
4551   tagacgatga tgactttcat acagatgtag aaacagcaaa agaattaaaa ttaaaagggt atacagttta
4621   attcactact aataaattca aagtatttgc tgcatttca gatactgtaa ttcattggatc agatgtagat
4691   ttagtaaaact ggggtgaaaa cgcactacaa tcagggtctag cagctaaaga acgtaaagat gccttagcag

```

4761 taagtccctaa atctggatta gatcacatgt cattttacaa tggatctgtt aaagaagtgt agggagcaga
 4831 catgtatgat gctattatta acgcttttagc agatttacat gaagattacc gtgataacgc aacaatttat
 4901 atgcatatgt cggattatgt caaaattatt agtgttcttt caaatggaac acaaatcttc ttgacacac
 4971 cagcagaaaa agtatttggc aaaccagtag tatttacaga tgcagcagtt aaacctattg tgggagattt
 5041 caattatttt ggaattaaact atgatggaac aacttatgac actgataaag atgttaaaaa aggcgaatat
 5111 ttgtttgtat taactgcatg gtatgatcag ccccaaaaagg ttaatgtaac agctaaggct aaatcagctg
 5181 aagaaaaatc aggttcatta cccagctaaag agtttagaag aaattaaatt gtggttgaga attgactata
 5251 taatatcagc cgaatagggtg tgaatgaaatg agtttagaag aaattaaatt gtggttgaga attgactata
 5321 atttcgaaaa tgatttaatt gaagggtctca ttcaatcggc taagtctgaa ttaattataa gtgggttctc
 5391 agattatgat aagatgact tggaaatccc gcttttttgt acagcgatta gatatacat tgcagagat
 5461 tatgaaagtc gtgggtactc aaatgaccaa tctagaagca aggtttttta tgaaaaagga ttgcaaaaaa
 5531 tgattctgaa attaaaaag tggtaggtga tttttaaatg gaatttaatt aatttaaga tcgcgcatat
 5601 ttttttcaat atgtaataa agggcctgat ccagatgaag aggaaaaaat gaagtgttat agttgctttt
 5671 gtaaaatata taactcttct atgaaagata gagaaatttt aaaaagcact gaattaaatt tgacagagc
 5741 cataattatg aggtcttcta aaattgaaata tctaccacaa acaaatcact tagttaaaat tgacagagc
 5811 ttatatctcg ataaaattat caacattaaa ttgatacacc agatattggc tataatacag
 5881 tgggttttat agaaaaatga gtgtagaat taaagggata cctgaagtgt tgaagaattt agaatcggta
 5951 tacggttaaac aatcaatgca agctaagagt gatagagctt taaatgaagc atctgaattt ttttaaaagg
 6021 ctttaagaa agaattcgag agttttaaag atacgggtgc tagcatagaa gaaatgacta aatctaagcc
 6091 ttatacaaaa ctgaggaaagtc aagaaagagc tgttttaatt gaattggtag gccctatgaa tcgcaaaaaa
 6161 attattcact tgaatgaaca tggttatata agagatgaa aaaaatatac accaagaggt tttggagtta
 6231 ttgcaaaaaa attagctgct aatgaacgga agtatagaga aattataaaa aaggagttgg ccagataaat
 6301 gaatatatta aacaccataa aagaaatttt attatctgat gcagagctcc aaacataatat aaattctaga
 6371 atatactatt ataaagtcac tgaaaatgct gaaacttcca aaccttttgt tgtattaca cctatttatg
 6441 atttaccttc agactctatg tctgataaat atcttagtga agaatactta attcaaatag atgtagaatc
 6511 ttcaataat cagaaaacaa ttgatataac aaaaacgaata agatattctgt tatatcaaca aaatttaatt
 6581 caagcatcta gtcagttaga tgcctatttt gaagaacta aacgttatgt gatgtcgaga cgttatcaag
 6651 gcataccaaa aaatatatat tataaaaaat agcgcatcga ataggtgtgc tttttaattt ttaaggagga
 6721 aataagcaat ggcagaagga caaggttctt ataaagtagg ttttaaaaga ttatacgttg gattttctaa
 6791 ccagaaagca acaaaagtag ttaaaagcat gacatgggaa gatgaaaaag gtggtacagt tgatctaatt
 6861 atcacaggtt tagcaccaga tttagtagat atgtttgcat ctaacaaagc tgtttggatg aaaaaacaag
 6931 gtactaatga agttaagctt gacatgagta tttttaatat tccaagtgaa gatctaaata cagttatttg
 7001 tctgttctaa gataaaaaat gtacatcttg ggtaggagag aatacaagag caccatacgt aacagttatt
 7071 ggaagatctg aagatgggtt aacaggtcaa ccaggtgacg ttgcgctact taaagggtact tttagcttgg
 7141 attcaattga atttaaaaca cgaggagaaa aagcagaagc accagagcca acaaaattaa ctggtgactg
 7211 gatgaacaga aaagttgatg ttgatggtag tccacaagggt attgtatagc ggtatcatga aggttaagaa
 7281 ggaagagcag aattcttcaa aaaagttatc gttggatata cggacagtgga agatcattca gaggattctg
 7351 caagttcgtt acccagctaa ccccaaaaat gttgaagtag cagtttaattc aaaaatctgca acagttctag
 7421 cagaatagggt gctttcaaaa taaatcaaaag gagaataatt tatgactaaa actttaagg ttttaaaagg
 7491 agacgagctc gttagcttctg aacaaggtga aggcanaagtg tcagtaactt tatctaattt agaagcggat
 7561 acaacttatc caaaaaggtac ttaccaagtg gcatgggaag aaaaatggtta agaatctagt aaagtgtatg
 7631 tacctcaatt caaaaacaaat ccaattctag tctcaggcgt atcatttaca cccgaaacta aatcaatcac
 7701 ggttaaatgct gatgacaatg ttgaaccaaa cattgcacca agtacagcaa cgaataaaac gttgaaatat
 7771 acaagtgaac atccagagtt ttgtactgtt gatgagagaa caggagcaat tcacgggtga gctgagggaa
 7841 cttcagttat cactgctacg tctactgacg gaagtgaaca gtctggacaa attacagtaa cagtaacaaa
 7911 tggataaata tttgagacgc agaataatct cgtctttttt atttgaataa aaggagctaa tacaatgatt
 7981 aaatttgaaa ttaagaccg taaacagaga aaaaacagaga gctatacaaa agaaagatgtg acaatgggag
 8051 aagcagaaaa atgctatgag tatttagaat tagtaaatca agagaataaa aaagaagtac ctaacgcaac
 8121 aaaaatgaga caaaaagagc gacagttatt agtagattta ttaaaagatg aaggattgac tgaagaagat
 8191 gttttgaaca agatgagcac taaaacttat acaaaaagcct tgaagatat atttcgagaa atcaatgggtg
 8261 aagatgaaga agattcagaa actgaaccag aagagatggg agcagtatgg gtggacatta actgaagtca
 8331 attttatcga acattaagaa ttctgtatgg ttctgtatgg agagataaaa gaagagactg aagaaaaaca
 8401 gaaaacagcc gtatgtaaaa ctttttagaaa tacttaatga agagaataaaa gctagaaaagg aggttaatat
 8471 aagtgaacaa aaagtcat ttagccttga gctgaaatta gaccttttag gtgtccaaga aggcataag
 8541 gaatgaaaaa gtagaaggca tgactcttga agtgaaatga aagctaatct gtcattcttt gataagctctg
 8611 ggttttaagc gacaaattag gtgtgtttaa ttaagggttt aaatgataag cttaaagttc aaaaaaagat
 8681 aaaaatcaat ggaaaagtat caggcgagaa agtttaacag atattatcaa aagctaaatc tagtgaataa
 8751 gtattctcaa gtagaagatg agcttaaaaa aggttaacgct aaaaagaaaa attagctctt gataaattca
 8821 gatgttgaga aagcatattt aaagctagta gaagctaaata aaaaagaaaa ataaacgta caaatcaacg
 8891 aagaagcctt aaaaatcttg aatacagaac gttgagagat gcagaacaaa agcttaagaa tagtaaccaa
 8961 taacaagat gcatacaaaa aacttaaaaa gacgcagtag agaaagcagc cyctaagcat aaagcacttg
 9031 gctactactg caaaactaaa aagagcaagt ttcaaaaact aaaaagtaca aatgataatc tttcaaatc
 9101 ttgaacaata taacaagaa ggaatcaag cttacgctaa aactaatact aaattaaagc aaacagaaaa agaattta
 9171 aaacgaaaaa atagaaaatt atagaaatg gctaatgctg caaaaagctga aacagctggtt aacaaagaaa
 9241 gatttaataa atactattaa gaactcatagc gctaatgctg tagataaagc ttcatccgaa atgaagactt
 9311 aagctgcttt aataatttta gagcgttcaa acttgcagtg caagcggatg tcatgtcaaa gaaatttagt
 9381 acaaatgata gctcaaaagc atttcggcaa cgtacagtag cgtatggcgt atctacacag attactttag
 9451 tctattggag ataaaatgac ttccctagga cgtacagtag cgtatggcgt atctacacag attactttag
 9521 ggttaggtgc agcattaaaa acaagtcag acttcgaag taatcaagcg gttgacttag gcgctaaaac aagtaaaagt
 9591 agcaagcag aaagacttaa aaagcatgtc tatggcagctt taggctttaa tgccaaacaa acaatggagg
 9661 gctaacgaag ttgctaaagg tatggaagaa ttggcagctt agaaatggct acaactgcaa ctgtaatggc
 9731 ctatgcccgg ttttatcagt gcagcagaag caagcggctg agaaatggct acaactgcaa ctgtaatggc
 9801 atcagcaatt aattctttcg gtttaaaagc atctgatgca aacctgtgtg ctgatttact tgcgagatca
 9871 gctaatgata gtgctgcaga tattcaatag atggggagat cattaaaaa tgcaggtact ccagcaaaag
 9941 cattaggagt tccaatagag gacacttctg cagcaattga agttttatct aactcaggtg tagaggggtc
 10011 tcaagcaggt actgcattaa gagcttctgt tattaggcta gctaattcaa gtaaaagtag agctaaggaa

10081 atgaaaaaat taggtattca ttgtctgat gctaaaggct aatttgttgg catgggtgaa ttgattagac
10151 agttccaaga caacatgaaa ggcacgacga gagaacaaaa actagcaaca gtggctacaa tagctggcac
10221 tgaagcagca agtggatttt tagccttgat tgaagcgggt ccagataaaa ttaatagcta tagcaaatca
10291 ttgaagaact ctaatgggtga aagtaaaaaa gcagctgatt tgatgaaaga caacctcaaa ggtgctctgg
10361 aacaattagg tggcgctttt gaactgtag caattgaagt tggtaaaagt ttaacgccta tgattagagc
10431 aggtgcgcaa ggattaaaca aattagttga tggatttaca catcttccctg gttgggttag aaaggcttcg
10501 gtaggcttag cgatttttgg tgcactcatt ggcctctgctg tctctgctgg tggcttatta acacgtgcag
10571 ttggaagcgc ggctaaaggc tatgcatcat taatatagac cattgctgaa aatacaatac tgcctaatac
10641 caattcaaaa gcaatgaaat ctttaggtct tcaaacctta tttcttgggt ctacaacagg aaaaacgcta
10711 aaaggcttta aaggattagc cggagctatg ttgtttaatt taaaacctat aaatgttttg aaaaattctg
10781 caaagctagc aattttaccg ttcaaaactt tgaaaaacgg tttaggatta gccgcaaat ccttatttgc
10851 agtaagtggg ggcgcaagat ttgctggtgt agccttaaaq tttttaacag gacctatagg tgcataaata
10921 actgctatta caattgcata taaagttttt aaaaccgcat atgatcgtgt ggaatgggtt aaaaacggta
10991 ttaacgggtt agggaaact ataaagtttt ttggtggcaa aattattggc ggtgctgcta ggaagctagg
11061 agagttttaa aattatcttg gaagtatagg caaaagcttc aaagaaaagt tttcaaaagg tatgaaagat
11131 ggttataaat ctttgagtga cgatgacctt ctgaaagtga gagtcaacaa gtttaaaagg tttatgcata
11201 ccatgggcac agcttctaaa aaagcatctg atactgtaaa agtgttgggg aaagggtttt caaaagaaac
11271 agaaaaagct ttgaaaaaat acgtacacta ttctgaagag aacaacagaa tcatggaaaa agtacgttta
11341 aaagcggttc aaataacaga agacaaagca aaaaaacttt tgaaaattga agcggattta tctaataacc
11411 ttatagctga aatagaaaaa agaaaataaa aggaactcga aaaaaactca gaacttattg ataagtatag
11481 tgcgttcgat gaacaagaaa agcaaaacat ttaactaga actaaagaaa aaaatgactt gcgaattaaa
11551 aaagagcaag aactcaatca gaaaatcaaa gaattgaaag aaaaagcttt aagtgtgggt cagattctag
11621 aaaaatgaaag aaaaagaaatt gaaaagcttg aaaaatcaaa acgtgacatc actgttaaaq aattgagtaa
11691 gactgaaaaa gacgaagagc gtattttagt aagaatgcaa agaaaacagaa atgcttattc aatagacgaa
11761 gcgagcaaaq caattaaaga agcagaaaaa gcaagaaaaa agaaagtgac aagcaaatatg
11831 aagatgatgt catgtctata aaaaataacg tcaacctttc taagtctgaa aaagataaat tattagctat
11901 tgctgatcaa agacataagg atgaagttaa aaaggcaaaa tctaataaag atgctgtagt agacgttgtt
11971 aaaaagcaaa ataaagatat tgataaagag atggatttat ccagtggctg tgtatataaa aatactgaaa
12041 agtgggtgaa tggccttaaa agttgggtgt ctaacttcag agaagaccaa aagaagaaaa gtgataagta
12111 cgcataagaa caagaagaaa cagctcgtag aaacagagaa aatataaaga aatgggttgg aaatgcttgg
12181 gacggcgtaa aaactaaaac tggcgaaagt tttagtaaaa tgggcagaaa tgctaactcat tttggcggcg
12251 aaatgaaaaa aatgtggagt ggaatcaaaq gaattccaaq caaattaaqt tcaggttggg gctcagccaa
12321 aagttctgta ggtatcacca ctaaggctat agctaatagt actggttaaa ggtttggaaa agcttggcaa
12391 tctgttaaat cgactacagg aagtatttac aatcaaaacta agcaaaagta ttcagatgcc tcagataaag
12461 cttggggcga ttcaaaactt atttggaaaq ggacatcaaa atggttttagc aatgcataata aaagtgcmaa
12531 gggctggcta acggatagtg ctaataaaat gcgctcgaaa tgggataata tttctagtac agcatggctg
12601 aatgcataaat ccgtttggaa aggaacatcg aaatgggtta gtaactcata caaatcttta aaagggttgg
12671 ctggagatat gtattcaaga gccccacgatc gttttgatgc aatttcaagt tggcatgggt ctaacgctaa
12741 atcagttatc aatgggttta gaaaatggct atcaagaaca tatgaatgga tttagatatg ttgtaaaagt
12811 atgggaagag ctggcgctga tttaggtaaa aatgttgcta ataaagctat tggcggttta aatagcatga
12881 ttggcggtat taataaaaaa tctaagacca ttactgataa aaatctcatc aagccaatc ctacattgtc
12951 tactgtgact tttagcgaaa aggggtgtag taccgataat tggggagcat taacgcaacc gacatttgct
13021 gtattaaatg atagaggttc tggaaacgcc ccaggtgggtg gaggtaacaa agtaattcac agggctgacg
13091 gaacattcca tgcaccccaa ggaagagatg tgggtgttcc actaggagtt ggagatagtg taataaatgc
13161 caactgacat ctgaagttac agcggatggg tgttttgcca aaattccatg ggggtacgaa aaagaaagat
13231 tggctagacc aacttaaaag taatataggt aaaaagcag gagaatttgg agctacagct aaaaacacag
13301 cgcataaatc caaaaaaggt gcagaagaaa tgggtgaagc agcagggcat aaaaacaaag atgggtgcac
13371 ttgggttaggc gatanaaatc gcgatgtgtg ggattacgta caacatccag ggaactagat aataaagta
13441 atgtcaggtt caaatattaa tttttggaggc ggaactaacg tacagtaaaa attgctaaag gcgcgtactc
13511 attgctcaaa aggaatattaa tagacaaagt aaaaatcgtg tttgaagatt ttgggtgggg agggcatggg
13581 agctatctat ttgaatatcc aatctggcaa agatttggac gctacacagg tggacttaac tttaatgacg
13651 gtgctcacta tgggtatagc tttggttatgc ctactggaa caaactttat gccgttaaaq gtgggtatgc
13721 agataaggta acgtactgatt acggtggcgg taattctata caaattaaqa ccggtgctaa cgaatggaa
13791 tggatatatg atttatctaa gcaattagca agacaaggcc aacgtattaa agctggctaa ctgataggga
13861 aatcaggtgc tacaggtaat ttctgttagag gacacacact acatttccaa ttgatgcaag ggtcacatcc
13931 agggaaatgat acagctaaaq atccagaaaa atgggtgaag tcaactaaaq gtatggcgct tcgaaggtgt
14001 tcaggtgtta ataaagctgc atctgcttgg gcaggcgata tacgtcgtgc agcaaaacga atgggtgtta
14071 atgttacttc ggggtgatga ggaatatca ttatgctgat tcaacacgaa tcaggaggaa atgacgggtat
14141 aactcaatct agttcgctta gagacatcaa cgttttacag ggcaatccag caaaaggatt gcttcaatat
14211 atcccaaaaa catttagaca ttatgctgtt agaggtcaca acaatatata tagtgggttac gatcagttat
14281 tagcgttctt taacaacaga tattggcgct cacagttaa ccaagagggt ggttggcttc caagtggctc
14351 aagaagatat gcgaatgggt gtttgattac aaagcatcaa cttgctgaag tgggtgaagg agataaaacg
14421 gagatgggtta tccctttaac tagacgtaaa cgagcaattc aattaaatga acaggttatg cgcacatcg
14491 gtatggatgg caagccaaat aacatcactg taataatga tacttctaca gttgaaaaat tgttgaacac
14561 aattggttatg ttaagtgata aaggaaataa attaacagat gcattgattc aaactgtttc tctcaggat
14631 aataaacttag gttctaatga tgcaattaga ggtttagaqa aaatattgtc aaaaacaaagt gggcatagag
14701 caaatgcata taatatatg ggagggttga ctaattaatg caatcttttg taaaatcat agatgggtac
14771 aagggaagag taataacaga ttttaattcag cttatatttt tagatgcaag ggtgaaagt ccaaacacca
14841 atgataacag tgcactattt aacggagtag atgggtattt accggggcca attagttttg cgcctttttc
14911 attagtatta agttttggct atgatgggtat agatgttata gatttaaat tattttgagc tattgttga
14981 tctgtgttta atcgacagca tcttattact gttattactt ctcaaatgcc tgggtgttaa tatgagtg
15051 atacagctaa tgttcatctt aattttaaag atgggtcttc aactgaaatt gaagtaagtt taaatgttta
15121 taaaaggtat tctgaatcag ttaattggac cgatagcgag tcttattcag actctaattg gatgtttgaa
15191 aatggaaatc ctcttgattt cacacctaaa tataactcata catcaaatca atttactatt tggaaacggt
15261 ctactgatag gatcaatcca cgattcaagc acgatttgaa aatattaatt aatttaaatg cgagtgagg
15331 atttgaactg gtttaactata caacagggtga tatttttaag tacaacaaaa gtatagataa aaacactgat

15401 tttgttttag atgggtgtgta tgcataatcga gatataaata gagtgggaat tgatacaaat agaggcatta
15471 taacatttagc gccaggtaaa aatgaattta agattaaagg agacatcagt gatattaaaa ctacatttaa
15541 gtttcctttt atttataagg aggtgattta atggattatc atgattcattt atcagtaatg gatttttaatg
15611 aattgatttg tgaaaattta ctatagtag attatgggtt ttttaaagaa tattatgaac tgaatgaagc
15681 taggtacatc acttttacag tttatagaac tactcataat agttttgttt tcgatttact aatttgtgaa
15751 aacttcataa tttatcatgg tgaaaatac acaattaagc agacagcgcc aaagggtgaa ggtgataaag
15821 tttttattga agttacggca tatcacataa tgtatgaatt tcaaaatcac tcagtggaaat caaataagct
15891 tgatgacgac agtagcgaaa ctggtaaaac gccagaatac tcttttagatg agtacttaag atatggattt
15961 gcaaatcaaa aaacttcggt caaaatgacc tataaataa tgggaatttt taagcgaaaa gtaccgtttg
16031 acgaattagg taacaaaaac gggttagaatt actgtaaaga agcggtagac ctatttggct gtataattta
16101 cccaaatgat acggagatat gtttttattc tcttgaaaca ttttatcaaa gaagcgagaa agtgattcga
16171 tatcaatata atactgatac tgtatctgca actgtcagta catttgaatt aagaacagct ataaaagttt
16241 ttggaaaaaa gtatacagct gaggaagaaga aaaattataa tcttattaga acaactgaca ttaaatattc
16311 aaatggtttt ataaaaagaag gtacttatcg tactcgcaaca attgggtcta aagctactat taactttgat
16381 tgcaagtatg gtaatgaaac agttagattt aacataaaaa agggctctca aggttgaata tataagttga
16451 ttttagacgg caagcaaat cttgttttgc taagtctggt cagctgtaaa caatagattt
16521 aatcaaaaaat attgataaag gcaagcagct ttcagaatg atatttttag gagaagaccc caaaaataga
16591 attgatatat cttcaaataa aaaagctaa ccttgtatgt atgttggaa cagctcttaa
16661 atttaattgc tgacaactca ggtcgcaatc aatacaaaagc aattgttgac tacgtcgagc atagtcaaaa
16731 gcagtttggg attcgatatg ctaatacgca acaaaatgaa gatattcgaaa cacaggataa gctgttagaa
16801 tttgcaaaaa agcaaataaa tgatactctc aagactgaat tagatgttaa ttatataggt tatgaaaaaa
16871 tagagccaaag agatagcgta ttttttggtc atgaattaat gggatataac actgaattaa aggttgttaa
16941 acttgatagg tcaatccatc ttgttaacgc aatagatgaa gtgtctttca gcaatgaaat aaaggatag
17011 gtacaaatc taacaagcgt taacagacga gttattgcac aagataatag atataactat caagcaaat
17081 gtataaatca tttatcacct agtactttga atttctcttt cgagacaatg gatattggga gtgtattaat
17151 ataattggcaa cagaagaagt taaaatcaaa gcgctacttg aaaacgataa acagtacttt ccagctacac
17221 attggaagc tataaatggg ataccttatg caggcagtag tgatattgat ggatttgcctc aagacggtat
17291 catttcggtg gatgataaaa ataaattaga taatttaaaa ataggcgaag caggaattat tcaaaatagc
17361 attgtacaga aatcccaaaa cggtaaatg tggaataaa cagttgacga tagtgggaaa ctctgtacag
17431 tgcattttta ttgaaagga aggtgcattt tggaaaattt gtatttaata aaggatttgg gagctttagc
17501 aggtcgagat tatagagta aggaataaca aaacttaca agaataagac aatttgcgct tggcttgaca
17571 acagagttta agttgcatca gaaagctaaa acaattcaac acttctgtga gcaaatttat tataatcgta
17641 gatcgcaagc agcagtaaac aaacttttac aaagtcacaa taacgcactt gttgtggcac cactgaataa
17711 cagtgctaat gagattgttc aagctcgagt taatgtaaac ggcgaacct ttgacacatt aaaagaacat
17781 ttgacgagtt gggaaaccca aactcaaat aataaagagg aaactataag agaattaaat aagaccaaac
17851 aagaaatttc tgatctgag tatcgttttg aacctgataa gcaagaattt ttatttgtga cagaacttgc
17921 acctcttaca aatgcagtaa tgcaatcttc ctggtttgat aatagaacag gcatagtata catgacaaa
17991 gctagaaata atggctatat gctaagtcgt ctaagaccta atggctcaatt tatagacagc tcattgattg
18061 taggtggggg tcatggtaca cataacggtt atagatatat tgatgatgag ttattgattt atagttttat
18131 cttaaatggt aataatgaga atacattagt tctgttcaag tatacgccta atgtggaaat tagctatggc
18201 aagtatggta tgcaagatgt atttacagga caccagaaa aacctcatc caccctgtc ataaatgaaa
18271 aagaaaaata aattctatac agaattgaga gacctagaag tcactgggaa cttgaaaact caatgaatta
18341 tatagagata agaagtttag acgatgttga taaaaatatt gataaagttt tgcataaaat cagtatccct
18411 atgagactaa caaacgaaac ccaacccaatg cagggtgtga cttttgtatga aaaatacttg tattggtata
18481 caggagacag taatccaaat aatagaaact atttaacggc tttcgattta gaaacaggag aagaagcgta
18551 tcaggttaat gctgactatg gtggaacact agattcattt cctggcgaat ttgcggaagc agaaggttg
18621 caaatatact atgacaaaag tagtggtaaa aaagcttga tgcaggtgt tactgtcggg ggtgatgaa
18691 atagaacaca tctgattttc atgatttggc aaagaggtat tttgaaata cttcactcaa gaggcgttcc
18761 ttttatcatg agtgacacag gtggtagagt taaaccttta ccaatgaggc ctgataaact taagaatctt
18831 gggatgttaa cagagccagg tctttactat ttatacactg atcatacagt tcaaatcgat gatttcccat
18901 taccaagaga atggcgtagt gcaggttggt tcttgggaagt taagccacca caaactggcg gtgatgtaat
18971 ttagatattg acggttaata gttatgcaag gaatatgatg acttttgaaa ggggtgcttc tgggaagact
19041 ggagacattt cggactggaa ttatgtgctt aaaaatagtg gtaaatggga gagagtacct tcattcatca
19111 ctaaaatgct agatattaac atagttagca tctcgtttta ttaactacg gatgatataa aacgttttac
19181 agattttcca actgaacgta aaggggtagc tgggtggaac ttatatgtag aagcttcaaa cacaggtggc
19251 tttgttcata ggctagttcg taatagttgt acagcatctg ctgagatact attgaaaaat tatgatagta
19321 aaacaagtct agggccatgg actttacacg aagggagaat tataagttta tgagtaattt agagaaatct
19391 gtactatata atttagaaaa cacagcgcat tatgaaaata tttcaaatct agatataact tttagaacag
19461 gagagagtga tcttctgttt cttcttttta atatcactaa aaataatcaa ccgttattat tgagtgaaga
19531 aatatcaaaa gcacgaatag cgtattcgagg taaaggagtc atggtagttg ctccactaga aatattagat
19601 ccattttaaag gtatttttaa atttcaatta cctaattgat taattaaacg agatggaagt tatcaagctc
19671 aagtttcggt tgcagaatta ggttaattcag acgtggtagt tgcgagaga actatcacat ttaacgttga
19741 aaaaagtttg ttttagcatga ttccatctga aacaaaatta cactatattg ttgaatttca ggaattgaa
19811 aaaactatta tggatcgtgc gaaagcaatg gacgaggcta taaaaaatgg tgaagattat gcgagctga
19881 ttgaaaaagc taagaaaaaa ggtctatcag atattcaaat agcaaaatct tcaagtagat atgaatttaa
19951 gcaacttgct aatagccata tctctgattt ggaaaaataa gcgcaagcat attcaagaac attcgagag
20021 caaaagcgat atatggatga gaaacatgaa gccctcaagc agtcagtgaa tagtgggtgt ttagtcacaa
20091 gtggttctac ttcaaatgg caaaaagcta agattactaa agatgtaggt aagataatgc agattactgg
20161 aattgatttt atcatcttcc aacaagaat actgtcgaat atttagtagt aacttcagat tacaagcgt
20231 aattatccaa gaggtgttag tactaacggt actgttgaat ttttagtagt aacttcagat tacaagcgt
20301 tgacttatcg accgaacggt acaaataaag tgtttgttaa aagaaaaaga gcgggttcat ggtctgagtg
20371 gtcagaataa gctattaatg attacaatac accttttgaa actgttcaaa gtgcccatac aaaagctaat
20441 atggccgaaa gtaacgctaa attatacgca gatgacaagt ttaataaaag gtattcgggt atttttgatg
20511 gaacagcaaa tgggtgtgggc tctacattgt acttaaatga gagtttagac caattttatt tattaatttt
20581 tttgtgggact tttccgggtg gtagctttac agagtttggc agtcttttg gagggaggaa gatttcattg
20651 aatccctcaa atcttccaga tgggtatgga aatgggtgag gtgtttatga gtttggatta actaaactca

20721 gtcgtacatc ttttaactata tcaaacgatg tctatttcga cttagggaagt caaagaggct ctgggtcgaa
20791 cgcaaataga gggacaatta acaaaattat agggagtga aataatgca aatattagtt aacaagcgta
20861 atgagataat ttcatacgct atcatctggg gctttgaaga aggtattgat attgaaaatt taccagaaaa
20931 tttctctcaa gtttttagac ctaaagcctt taaatattca aatggggaaa tagtttttaa cgaagattat
21001 tcagaagaaa aagatgactt gcatcaacag attgacagt aagaacaaaa cacagtcgct tctgatgaca
21071 tcttacgaaa aatgggtgct agtatgcaga aacaagttgt tcaaatgaca aagttatcga tgcaggttaa
21141 taagcaaaat gcactaatgg caaaacaact tgtgacactt aataaaaaat tagaagagggt taaaggagag
21211 actgaaaaatg cttaaattaa tttcaccacac attcgaagat attaaaaat ggtatcaatt gaaagaatat
21281 agtaagaag atatacgctg gtatgtatgt atcggaagtt tagataaaga ggaatatgca attattacag
21351 gagaaaaagta tccagaaaat cttagagtcac aggttataat cttatggctt ttttaatttga ataaagtggtg
21421 tgggtgaatg tttggattta ccaaacgaca cgaacaagat tggcgtttaa cgcgattaga agaaaatgat
21491 aagactatgt ttgaaaaatt cgacagaata gaagacagtc tgagaacgca agaaaaaatt tatgacaagt
21561 tagatagaaa tttcgaagaa ctaaggcgctg acaagaaga agatgaaaa aataaagaga aaaaatgctaa
21631 aaatattaga gactcaaga tgtggattct aggtataata gggacgattc taagtacatt tgttatagcc
21701 ttgttaaaaa ctatttttgg catttaaaagg aggtgattac catgcttaag ggaattttag gatattagctt
21771 ttgggtcgtgt tttcgtgtta gtaagtgtaa gtaatagtt agagtcagtg cttcggcact ggctttttat
21841 tttggaaaaa aggagcaaac aaatggatgc aaaagtata acaagataca tctgtattgat cttagcatta
21911 gtaaatcaat tcttagcgaa caaagggtatt agcccagatt cagtagacga tgagaatata tcatcaataa
21981 tacttactgt tgttgcttta tatactactgt ataagacaa tccaacatct caagaaggta aatgggcaaa
22051 tcaaaaagcta aagaaatata aagctgaaaa caagtataga aaagcaacag ggaagcgcc aattaaagaa
22121 gtaatgcacac ctacgaatat gaacgacaca aatgatttag ggttaggtgtt gaccaatggt gatatacaaa
22191 aaccaagcag aaaaatgggt tgataattca tttaggaagc agttcaatcc tgatttgttt tatggatttc
22261 agtgttacga ttacgcaaat atgtttttta tgatagcaac agggcaaaagg ttacaagggt tatagcctta
22331 taattattcca ttgataata aagcaaggat tgaaaaatc gggcaaataa ttaaaaaacta tgatagcttt
22401 ttaccgcaaa agttggacat tgcgttttc cgttcaaaagt atgggtggcg agctggacat gttgaaattg
22471 ttgagagcgc taactcaaac actttcacat cgtttggcca aaattggaaat ggtaaagggt ggacaattgg
22541 cgttgccgaa cctgggtggg gtcccgaaac cgttacaga catgttcat attacagta cccaatgtat
22611 tttattagat taaaattccc agataaagta agtgtggag atcaagctaa aagcgttatt aagcaagcaa
22681 ctgcaaaaaa caagcagta attaaacctt aaaaaattat gctttagacc ggtcatgggt ataacgatcc
22751 tggagcagta ggaacggaa caaacgaaac cgttttata cgttaataa taacgcgcaa tatcgctaa
22821 tatttaagac atgcggctca tgaagtgcga ttatatgggt gctcaagtca atcacaagac atgatatca
22891 atacagcata cgtgtttaat tagagtaata aaaaagatta tggcttatat tgggttaaat cacaggggt
22961 tgacattgct ctagaataac atttagacgc agcaggagaa agcgcaagtg tggggcatgt tattattcca
23031 agtcaattca atgcagatac tattgataaa agtatcaag atgttattaa aaataactta ggacaataa
23101 gaggtgtaac acctcgtaac gatttactaa atgttaacgt atcagcagaa ataaatataa attatcgctt
23171 atctgaatta ggttttata ctaataaaaa tgatatggat tggattaaaga aaaactatga cttgtattct
23241 aaattaatag ccggtgcgat tcatgttaag cctatcggtg gtgtgatac tagtgaggtt aaaaacccag
23311 ttaaaaacga aagaatccg ccagtgccag caggttatat acccgataaa aataatgtac cgtataaaaa
23381 agaaactggt tattacacag ttgccaatgt taaaggtaat aacgtaaggg acggctattc aactaattca
23451 agaattactg gtgtattacc taataacgca acaatcaaat atgacggcgc atattgtatc aatggctata
23521 gatggattac ttatatgtct aatagtgga aacgtgttca tattgtcata ggagaggtag acaaggcagg
23591 taatagaata agcagttttg gtaagtttag tgcagtttga taattgtata tgatgaatct taggcaggta
23661 cttcgggtact tgcctattat ttaaaattaa taacacagta atttttacat gaataattta aatttttaaa
23731 aaacaaacgt ttttagtata taaattattt tgtgttcgta ttgtgtgcta tgattaaaaa gttgttatgg
23801 tcaactatct cgtgggttta tgtttattat caatcaaaat ataaattatt tataatttgt ttggtaatga
23871 accgggttttt ttcgaaataa tagtaaaaaa acacatttgt agatatttta aactcggtaa atcttttaat
23941 aaattatttaa ttttattaaa agttaaaaag gtttaataa aaaaatgtaac aaaaattata aagaaggaa
24011 atgattttta tgggtcaaaa aagactatta gctgcaacat tgcgtttagg aataatcact cttattgcta
24081 cttcgtttca tgaatctaaa gctgataaca atatttgagaa tattggtgat ggcgttgagg tagtcaaaag
24151 aacagaagat acaagtagcg ataatgggg ggtcacacaa aatattcagt ttgattttgt taaagataaa
24221 aagatataca aagacgcttt gattttaaaa atgcaagggt ttatcaattc aaagactact tattacaatt
24291 acaaaaaacac agatcatata aaagcaatga ggtggccttt ccaatcaaat attgggtctca aaacaatga
24361 ccccaatgta gatttaataa attatctacc taaaaataaa atagattcag taaatgttag tcaaacatta
24431 ggttataaca taggtggtaa ttttaaatgt ggtccatcaa caggaggtaa tgggtcattt aattattcaa
24501 aaaaatttag ttataataaa ataaaaagta ggtgataaga tgactcaatt tctaggggcy cttcttcta
24571 caggagtttt aggttacata ccatataaat atctaaacat gataggttta gttagtgaaa aaaaacagg
24641 tatcaatatt cctgtattat tgattttttc tattgaaaca tgtttgatat ggttttatag ttttatatt
24711 ttttaaatg ttgatttaaa aaatttgaat ttaattcagt tgaattcagg tgcctacagg tctaaaagca aatattttgt
24781 tttctattat ttttgtttta acagtggttg tatttaaccc ttttaattgt aaatttatta tctgggtta
24851 taatataacc agaaagtta tgaatttga ttgtataagc ttattagaca aaagagacaa gttgtttaat
24921 aacaacggtta aaccagtatt tatagttata aaagactttg aaaaacagaa cattgaagag ggtgaactta
24991 aaacctataa ttcagctggt agcagtttcg atttactaga agttgagcga caagatttca aagtatctga
25061 tttaccgtca aacagtgaa tgtatattaa acatacactt gttagacctt aacaacaaat taaattggat
25131 ttatatttaa tgaatgaata ctaattcttt tttctagctt tttctgataa agtgcctttt aatttttcgc
25201 tggcgcccg cttttcaaaa cttttgttta ttgggttact acgagtagct tctgtttttt tgtttttatc
25271 cgccataaaa tttccaccac cattcaacgt ctacacttgt aggcgttttt ttatttagta aagtcataat
25341 gaattcttct tgggttaact atctccactt attttttgtg aaataaattc caagttatta cgcgcattat
25411 gtgacgataa atctttaggt aactcataag tgaatgggtg attaccacta gttaaaaact catatactat
25481 agtttctttt tttatttgc aattagttat tttcattata aacttctttt caaacactgc tgaattagac
25551 gttctttata ttaaagcgcc acacagggcg ttgtaattcac aatacaactt tgcccattac ttttaattat
25621 ctaaacgaaag cgactttgat atcatcaca ctaatacaac gagtgcaatt gtaccatctt taatagaaat
25691 tatctacacg cttgataaga cttactccat ttttaacata ggttccattg aatcaccatt aactaaaaa
25761 tttctttcta ataaaagcgt atgttctctg ttttaacata ggttccattg aatcaccatt aactaaaaa
25831 caaaaactag catttgatgg cgtttcgtct tctttaaaaa atacttcttc atgcaaatat tcatcatata
25901 attcttctcc tatgcccaga ccagttgcac cacatgcaat atacgatact agtttagact ctttatatcc
25971 atctatagaa gtgactttat tctgtctctc caattgttca ttcgcatagt taagtacggt tcttggcggt

26041 ggagggtgtga gtttgtgtga tatggaagt atgtcggtat cgtctttgta tgtagtattt gattcactat
26111 acaaatcatt aatcttcaca ttgaagtact cagccaaaaat tttggcagtt gataatcgag gtctctcctt
26181 ttcatttttcc catcttgata tcttgccctt cgttaatttc attaatgctgg gatatttatt attaatgata
26251 gttgtctaatt gttccatagt catattttta ttttttctt agcttcttta aaccttcacc aatacccata
26321 cgaaaccctc cttatataag ataatttcat tataaaagt tcgaaaaaag aacgcaagga aatattattt
26391 gcaaaagtgt ttgacatcga aacttttatg atgtattctt aaatcaaggt gttacaaacy aaacaaaagg
26461 aggggggttca atgacaacta gtgtagcaga taaaccatct tcaaaaataa aaagcttgat tgcacttaaa
26531 ggaactaacc aaaaagaagt tgctaaagca atcggaatga gtagaagttt attgagtata aagataaatc
26601 gaattaatgg cagagatttt acaacttcag aagctaaaaa attagcagat cttttaaatg ttaagttga
26671 tgattttttt taaaactttaa gtttcgaaag tgacaactaa ataaaaataa ggaggacact atggaacaaa
26741 taacgttaac caaagaagag ttgaaagaaa ttatagcgaa agaagttaga aatgctataa aaggcgagaa
26811 accaatcagc tcaggtgcga ttttcagtaa agtaagaatc aataatgacg atttagaaga aatcaataaa
26881 aaactcaatt tcgcaaaaaga tttgtcgcta ggaagattga ggaagctcaa tcatccgatt ccgctaaaaa
26951 agtatcagca tggcttcgaa tcaattctat aaaaagctta tgtacaagat gttcatgacc atattagaaa
27021 attaaactta tcaatttttg gagtgacact taattcagac ttgagtgaat gtgaatacaa cctagcagca
27091 aaaaatttata tgaactttaa aaactattat ttatatatct atgaaaagag agtttcagaa ttaactatcg
27161 atgatttcga atgaaggagg aactacaaat gaaactacta agaaggttat tcaataaaaa acacgaaaaa
27231 tcaattgacg tgtggcatgg aaatcaatgg ttaaaagtga aagaagcaa attaaaaaa tataaagtgg
27301 tctcggatag agaaggttag aatatcttaa ttaaaatagc gcacttaatt agtgcaagta atcaagtgcg
27371 ctactgcctt acaatcctaa atcttttctg ctttttctt cttctgttaa tcccaataac acagaagagt
27441 aaatgctgaa atagtcacga gcaacgctat ctttagcgaa tgcattacg tcatcccgat ctcttgcga
27511 tctcgttatga atcttatgtc tatctagagc tcttaggtaat agcgagattg taatctcgta agcaatttct
27581 tctaaatcca taaatttctac ctctctccac tgggagataa ctaaatata taacaaaaa acttaaggga
27651 ggaacgacaa atgcaagctc aaaaacaaaa agtcatctat tactactatg acgaagaagg taataggcga
27721 ccattataga tccaatttaa tgacggatat gaaactgagg tccgattctca ttctacaaac aacaccattg
27791 aagaataacc atacgttaat aataacttat atgctctggg tgatgggtat gaatttaagt tagattgaat
27861 ttttgagaaa gatattgaaa agctaatttc cccataagat taagagacat actggatgtt ttgttaacga
27931 ctcttctaac tctgttccaa gttttatgt ctctaatatc atcgagaaat tcatggccag accaagtgtg
28001 gtcacataa atccaagaaa cgacctgccc ttcgatgaat ttcagatcgc aacaaataaa tttagcttct
28071 tctaatttta aaagttagta cattactgtt tcaaaatcac atttatcaa aataatatta tctgtgaaat
28141 tatgtcgatg aagtggttca cctattttct tatttagattc tatttctaa agcaagagtc taacgcaatc
28211 gtgattaagt ttcactctat cactctcata acaggagatc agcagaagg atcataaaca tcttaaggga
28281 aggaataaca aatgaacatt caagaagcaa ctaagatagc tacaanaaat cttgtctcta tgacacggaa
28351 agattggaaa gaaagtcatc gaactaagat attaccaaca aatgatagtt ttttacaatg catcatttca
28421 aatagcgatg ggacaaacct tatcagatat tggcaacctt cagccagatga cctcatggga aatgattggg
28491 aagttataaa cccaactaga gaccaggaat tattgaagca attttagaaa tgcataaat gatacttttt
28561 aaattgtttt tcaactcatt ttcaagctaa acaacagctc tgcctgaaat tgttcatatg taatagtgt
28631 tactagcata cagccgcttt aggaaccag agtttttaag tttatttaa tctgtattta catcttcgaa
28701 atgtagtttt tgaataact tttgtatgt atcttttagca cttccaaaat tattgcaggt taatttaacc
28771 gaaccttaact ttacacattc taataaatct ttgtagagta cggacaagat atattgttgg tctttagtaa
28841 gtgtatcaaa ttcactcagat atcaagggca tgttatcacc tctctagggt gataacaaaca ttatacaga
28911 aaggagcata aacaaatgaa cacaagatca gaagattgc gtataggcgt cccacaagtt tctagcaaa
28981 tctagtcttc ttcactctat ttaacggaaa aggaacgtaa cttaggagcg gaaattattg agcttattaa
29051 aaaaagttag tacagctact tagaataaaa caaagttttc tatgcattag atagagaact tcaatcaggy
29121 gcgaataaca acaacttcta acatttctc aaaggagtg tagagatgcc aaaaatcata ataccacaa
29191 caccagaaaa cacatatcga ggcgaagaaa aatttgtgaa aaagttatc gcaaaccta cacaatcca
29261 tcaattgttt ggagtatgta gaagtacagt atacaactgg ttgaaatatt accgtgaaga taatttaggt
29331 tagaanaaat tacaactgta tttattagca acgggaacat tgattaatat ttctaattta gaagagatt
29401 tgatcagaaa gcataaaaaa tgggtattagg aggtattatc aatgagcgac acatataaaa gctacctatt
29471 agcagtggtt tgcctcaggg tcttagcgat tgtactcatg ccgtttctat acttctactac agcatgttca
29541 attcggggat tcgcaagtat cggcaacttc atattttata aggaatactt ttatgaagaa taagaanaat
29611 gctacttgtt ggagcaagta acagtgcag atgagcaat gtcttaata attatataag gaggttattaa
29681 tatgacctta caacaaaaaa tactatcaca ttttgcaaca tatgacaatt tcaattctga tgaattgtt
29751 gaagtttttg ggtatcttaa aacacatgca aaatccacac tttcaagat taagaaaaaa ggaaagattg
29821 aattggaaag ttgggtatc tggcgtgtg ttgaaccgca gttacattta actgtgttag aacgttaagaa
29891 agagatatta gaagaacaat tcgagttatt ggcgaagata aacyaacaat gtgatgacct tagagaaata
29961 gaagaacgca tcaagttaat gattcgttta gccaaacca ttttaaggag agttaatcaa tggcaatatt
30031 agaaggtatt tttgaagaat taaaactatt aataagaat ttacgtgtgc taaatactga actatcaact
30101 gtatgttcat caattgtaca agagaaagt aaagaagcac caatgctaaa agatgaana gctcaactgg
30171 aatcagttga agaagtttag gaaacttctg ctgatttaac taaagattat gttttatcag taggaaaaga
30241 gttccttaaa aaagcagata cttctgataa gaaagaattt agaaataaac ttaacgaact tgggtcggat
30311 aagctatcta ctatcaaga agagcattat gaaaaaattg ttgattttat gaatgcgaga ataaatgcat
30381 gaagctagat cactcaata gagctcatgc aaagcttagt gcaagtgag caaaaaatg gctaaactgt
30451 ccaccagata ttaaggcaag tgaaggtatt cgagataaaa gttcagtttt tgcgtaagaa ggtacattcg
30521 ctcctagatt aagttagtta tatttcagtc ttaaatatga aggcctaaca cagtttgagt ttaataaagc
30591 ttttcaaat tataagcga atcaatata cagtgaagag ttgcgcgaat atgttgaaga gtactgagct
30661 aatgtagaag aaaaatataa cgaagctttg agtagagatg acgatgtaat agctttattt gaaacaaaat
30731 tggatttagg taaatacgtc cctgaatctt ttggtactgg tgatgtcatt atattttcag gtggtgact
30801 tgaatttatt gaccttaag acggttaagg cattgaagtt tcatgctatg ataatctca acttagatta
30871 tatggcttgg gcgcataatg actgcttagt ttaagttagt acattcatat agttcgcag actatcatat
30941 aaccacgaat agataacttt tctactgaag agttaccaat atcaagatta cttcaatggg gaaccgattt
31011 ttgttaacca tataccagac tggcttataa cgttgaaagt gagtttaag caggtagcca ttgtagattc
31081 tgaagataa agcatttcag tagaacagct gcagaatata tgaanaatgt gcctcaaaag ccaccacatt
31151 tgttgagtga tgaagagatt gcagaacttt tatataaact gctgacatc aaaaatggg ctgatgaagt
31221 agaaaaatat gcaactagatc aagcgaaaga aaatgataa aactattctg gttggaagct tgaagaagt
31291 cgctcgcga gaatgataac tgatacaaat gcaacgcttg aaaaagttagt tgaagcaggt tataaacctg

31361 aagatattac agaaaccaag ttacttagca ttacgaattt agaaaaatta atcggcaaaa aagcattttc
31431 taaaattgca gaaggtctta tagaaaagcc acaaggtaaa ttaacacctg ctaccgagtc tgataaacga
31501 ccagctataa agcaatctgc tgaagatgat ttgacaaac tataaaaatt aaaaaggacg gtatataaac
31571 atgaaagcaa aagtattaaa taaaactaaa gtgattacag gaaaagtaag agcatcatat gcacatattt
31641 ttgaacctca cagtatgcaa gaagggcaag aagcaaaagta ttaactcagt ttaactcatt ctaaatcaga
31711 tacaagtacg ataaaagcca ttgaacaagc tatagaagct gctaaagaag aaggaaaagt tagtaagttt
31781 ggaggcaaaag ttctctgcaa tctgaaactt ccattacgtg atggagatac tgaaaagagaa gatgatgtga
31851 attatcaaga cgcttatctt attaacgcac caagcaaaac agcacctggt attattgacc aaaaacaaaat
31921 tagattaacg gattctggaa ctattgtaag tgggtgactat attagagctt caatcaattt atttccattc
31991 aacacaaatg graataaggg tatcgcagtt ggattgaaca acattcaact tgtagaaaaa ggcgaaacct
32061 ttggcggtgc aagtgcagca gaagatgatt tcgatgaatt agacactgat gatgaggatt tcttataagt
32131 caatagggtg gggttttagc cccactttaa ttttaagaa attgaggtgt caagaatttg aaattttaga
32201 atatagatat tgaaacatat agcagtaacg atatttcgaa atgtggtgtc tataaataca cagaagctga
32271 agatttcgaa atcttaatta tagcttattc aatagatggt ggaccgatta gtgagattga catgactaaa
32341 gtgataatg agcctttcca cgctgattat gagacgttta aaattgctct atttgacctt ataccctgta
32411 agtatgcatt caatgctaatt ttgaaagaa cttgtcttgc taaacatttt aataaacaga tgccacctga
32481 aagaatggatt tgcacaatgg ttaattcaat gcgtattggc ttacctgctt cgcttgataa agttggagaa
32551 gttttaagac tacaaaacca aaaagataaa gcaggtaaaa atttaattcg ttatttctct ataccctgta
32621 agccaacaaa agttaatgga ggaagaacaa gaaatttggc tgaacatgat cttgaaaaat ggcaacaatt
32691 tatagattac tgaattcgag atgtagaagt agaaatgaca attgctaata aaattaaaga ctttccagta
32761 acgttaattg aacaagcata ttgggttttt gaccaacata taaacgacag aggtattaaag ctttctaatt
32831 cattgatgtt agggagctaat gtgctcgata agcagagtaa agaagaattg cttaaacaag ctaaacatat
32901 aacaggttta gaaaactcta atagtcctac acagttattg gcttggttaa aggtagaaca aggattagat
32971 atacttaatt tacaagaaga aacggttcag gattacttaa aagtagcaac aggaaaagct aaaaaaattg
33041 tagaaattag attgcaaatg tctaaaacca gtgtgaaaaa atacaacaaa atgcatgaca tgatgtgcag
33111 tgatgaacgg gtaagaggtc tgtttcaatt ctacggtgcc ggtactggaa gatgggcagg tagaggtgta
33181 atacttcaga atttaacaaa gcatattatt tcagatactg aattagaatt agcaagagat cttatctaaag
33251 aacaacgttt tgacgattta gatttattac tcaatgttca tctcgaagac ttattaagtc aattagttag
33321 gacgacattt actgctgaag aaggttaatga actagcagta agtgattttt ctgcaataga ggcaagagtc
33391 atagcatggt atgcaaaaga acaatggcgt tttagatgtg tcaacacaca cggaaaagata tatgaagcat
33461 cggcttctca aatgtttaat gtaccggtag aaagcataac taaaggcgac cctctcagac aaaaaggaaa
33531 agtgcctgaa tttagctttag gctatcaagg tggcgctgga gctttaaaag caatgggtgc attggaaatg
33601 ggcattgaaag aaaaacgagtt acaaggttta gttgatagt ggcgtaacgc aaactcctaac atagttaatt
33671 tttggaaggc ttgccaagag gctgcaatta atactgtaaa atccgaaaag acgcatcata cacatggact
33741 tagattttat atgaaaaaag gttttctaatt gattgaactg cctagtggaa gagcttttagc ttatccaaaa
33811 gcttttagtt tgtaaaatag ttggggtagt caagttgtttg aatttatggg gttagatctt aaccgtaaat
33881 ggtcaaagtt aaaaacgtat ggtgggaagt tagtcgagaa tattgttcaa gcaactgcaa gggatttact
33951 tgcgatttct atagcaagggc ttgaagcatt aggtttttaa atagttggcc atgtccatga tgaagtaatt
34021 gtagaataac ctgagaggtc aaatggactt aaggaaatcg aaactatcat gaataagcct gttgattggg
34091 caaaaggatt gaatttgaat agtgacgggt ttacttctcc gttttatatg aaggattagg agtgbgatg
34161 catgcaacat caagcttata tcaatgcttc tgttgacatt agaattccta cagaagtcga aagtgttaat
34231 tacaatcaga ttgataaaga aaaaagaaat ttggcgact atttatttaa taactccaggt gaactattaa
34301 aatatcaacgt tataaatatt aaggttttag atttagaggt ggaatgatgg ctagaagaaa agttataaga
34371 gtgcgtatca aaggaanaact aatgacattg agagaagttt cagaaaaata tcacatatct ccgaacttc
34441 ttgatattag atacaaacat aaaaatgcgc gcatgaatt attgtgtgga agaaaagact caaaacttaa
34511 agatgaagtt gaatatatgc agagtcaaat aaaagatgaa gaaaaagaga gagaaaaaat cagaaaaaaa
34581 gcgattttga acctatacca acgaatgtg agagcggaat atgaagaaga aagaagaga agattgagac
34651 catggcttta ttgatggaac ccacaaaac attcacgtga tccgtactgg ttcgatgtca cttataacca
34721 aatgttcaag aatggagtg aagcataatg agcgtaatca gtaacagaaa agtagatatg aacgaagcgc
34791 aagacaatgt taagcaacca gcgcactaca catacggcga cattgaaatt atagatttta cgaacaggt
34861 tacggcacag tatccacctc aactagcat cgcaataggt aatgcaataa aatatttgtc tagagacct
34931 ttaagaagtg gtcgatgaga tttagcaaa gcaagtttt acgtccaaag agcttttgac ttgtgggagt
35001 gatgaccatg acagatagcg catgtaaaag atacttaaac caatttttcg gatcgaagag atactctgat
35071 caggataacg aacgagtggt acatatccat gtagtgaatg gcacttatta ctttcacggg catatcgtac
35141 caggctggca agcggtgaaa aagacatttg atacagcgga agagctcgaa acatatata agcaacatgg
35211 tttggaatc gaggaacaga agcaactaac tttattttaa ggagatagaa atgatgaaaa tcaagttga
35281 aaaaataatg aaaaatagac aattaattaa gtgggcgcga gaaaatccgg agctattcatt tggcagaaaa
35351 tattatacaa cagacaaaaa tgatgaaaac tttatttact tcggtgtttt taaaaattgt tttaaaaa
35421 gcgattttat attagttaat gctactttta gtgtcaaggt tgaagaagaa gtaaccgaag aaactaagtt
35491 tgataggttg tttgaagtgt acgagattca agaaggagtc tataaatctg catcatatga gaatgctagt
35561 ataaacgaac gtttaaaaaa tgacagaatt tttcttgcga aagcatttca catcttaaac gacgacctaa
35631 ctatgacgtt aattttgaaa gaaggagagt tgattaaata atggaacacg gttcaaaaaga atattacgaa
35701 aagcaaatgt aatactggtt tgatgaagca agcaagtttt tgaagcaacg tgatgagctt attggagata
35771 tagctaaagt aagagagtg c aacaaagagc tgggaagaa agcaagtga tgggtaggtt attgcaagag
35841 cggttgatga aacttaataa acgaatttgg caaagatggt gaaagagtta aatttggaaat ggaattaaac
35911 aataaaaatt ttaaggagga agacgcaaat gaataaccgc gaacaaatcg aacaatcagt tattagtgt
35981 agcgcgata acggcaatga ccagagggga ttattaaaag agattgagga cgtgtataag aaagcgcaag
36051 cggttgatga aactactgag ggtttacctg atgctatgca agatgcaatc aaagaagata ttggtcttga
36121 tgaagcagta ggaattatga cgggtcaagt tgtctataaa tatgaggagg agcaggaana tgactaacat
36191 attacaagtg aaactattat caaaagacgc tagaatgcca gaacgaatc ataagacgga tgcaggttat
36261 gacattattt cagctaaaac tgcgtactt gagccacaag aaaaagcagt gaccaaaaac gatgtagctg
36331 taagcattcc agagggctat gtcggtttat taactagccg tagtgggtga agtagtaaaa cgcatttagt
36401 gattgaaaca ggaagatag acgcgggata tcatggtaat ttagggttta atatcaagaa tgataatgaa
36471 acgttgatga gtaggatat gagtaacttt ggtcgaggtc cttctggtat agatggaaaa tacacctac
36541 tacctgtaac agataaaatt ttatgtatga atggtagtta tgcataaat aaaggcgaca aactagctca
36611 attggttatc gtgcctatat ggacacctga actaaagcaa gtggaggaat tcgagagtgt ttcagaacgt

36681 ggagcaaaag gcttcggaag tagcggagtg taaagacata ttagatcgag tcaaggaggt ttcggggaag
 36751 tgagtgacat gctagaaata ttttctcatg ggtttcggtg ttatctatct tgcgcacatag gtatctatct
 36821 tctcaagagt aaaaagacta tacacacaaa cctatatgaa atgttgttga ttgtactatc ctttgtgaca
 36891 tctacatttg ctgataaaca tcaaaagacg catatcttaa tagcattttt agtaattgtt tttatgagta
 36961 agctcaaaac agttcaaggg agctatgagg aatgacacaa tacctagtca caacatttaa agattcaaca
 37031 ggacgttaagc atacacacat aactaaagct aagagcaatc aaaggtttac agttgttgat ggggagagta
 37101 aagaagaagc gaaagagaag tacagggcac aagttaaaag aaatgcagct attaaattag ggcaggtgtt
 37171 tgaataataa agggagtggt ggaatgact aaacaaatac taagattatt attctacta gcgatgtatg
 37241 agctaggcaa gtatgtaact gagcaagtat atattatgat gacggcta at gatgatgcag aggcgcgag
 37311 tgactttgaa aaaatcagag ctgaagtttc atggtaatag ctattatcat ttttgaatta attatattaa
 37381 tgtgttttagc aatagcactg gaggtgttgt aaatatgtgg attgtcattt caattgtttt atctatattt
 37451 ttattgatct tgttaagttag catcttctcat aagatgaaaa ccatagaagc attcggagat atgaatcgtt
 37521 atcttttcaa gcagtrtagta aaaaaaatg gtgttgaagg tatagaagat tatgaaaatg aagttgaacg
 37591 aattagaaaa agatttssaa gctaaagaga ggcgtttggt tctctgttct atttssaaata atgaaaggag
 37661 ccgaacatgt tagacaaagt cactcaaat aacacaaata aatatgatcg tgatgtttca tattcttatg
 37731 ctgctagtcg tttatctaca cattggacta atcacaaat ggcttggctt gactttatgc agaagctagc
 37801 acaaacagtg aagactaaag aagatttaac tgagtacaat aaaaatgtct agtctgaaca agcagatata
 37871 aaagatgttg ggggatttgt cgytggttat ttaaaagaag gcaaacgacg tgcgtgtcaa gtcargaatc
 37941 gttcaatggt aacacttgat atcgattatg ctgctcaaga tatgactgac atattatcta tgtttatga
 38011 ttttgcatac tgtttatatt caacacataa gcatagagag ataagctcaa gactgcgttt agtgattcct
 38081 ttaaaacgaa atgtaaatgc agatgagtat gaagctattg ggcgttaaag cgcagatatt gttggcatgg
 38151 attacttcca tgatacaact tatcaaccac ataggtta at gtattggcct tcaactagta acgatgcgga
 38221 atttttcttt accatgaag atttacctt gttagacca gataaaatat taaatgaata tgttgattgg
 38291 actgacacat tagaatggcc aacgtcttca agggaaagaga gtaagactaa aagattagca gataagcaag
 38361 ggcaccagaa agaaaagcgg ggaattgttg gtgcattttg tagagcctat acgatagaag aagctataga
 38431 aacttttatt cctgatttat acgaaaaaca ttctactaac cgttatacct atcatgaagg ttaactgcga
 38501 ggtggatttg tgttatcaga aaataacaag tttgcttatt ctcatcataa tacggatccc gtaagcggta
 38571 tgccttgtaa cagttttgat tttagtcgca tacacttata tgggtgtcaa gatgaagacg ctaaaaacga
 38641 tactccggtt aatcgactac ctagtataaa agcaatgcag caaagagcgc aaaaatgata agttgttaaa
 38711 aagcaattaa ttaacgacaa aatgtctgat gcaatgcagg atttcgata aatagtaaat agcgtatgat
 38781 catgggtcga gacgttagaa attacttcga aaggtacttt caaagctagt atcccaata tagaatttat
 38851 attgctgaat gatccaaatt taaaaggaaa aatagcattt aatgaattta caaaacaaat tgaatgttta
 38921 gggaaaatgc catggaataa taattttaa atacgtcaat ggcagaagcgg tgatgatagc agtttaagaa
 38991 gttatatcga aaagatttat gacatacacc attcaggcaa aacaaaagat gccattataa gcgtagcaat
 39061 gcaaaatgcc tatcatccag taagagatta tctaaataaa atatcgtggg atggacataa acgtcttgaa
 39131 agcttatatta tcaaatactt aggtgttgaa gacactgaag tgaatagaac aactaccaaa aagcgattga
 39201 ctgctggaat cgtctgagta atggagccag gatgtaaatt tgacratatg cttacacttt atggctctca
 39271 aggtgtaggt aaatctgctt tgctaaaaaa aataggttgt gcatgttttt ctgacagttt agtttctgtt
 39341 actggtgaag agacatatga ggcattacaa ggogtttgtt taatggaaat ggcagaactt gcagctacaa
 39411 gaaaagctga agttgaagct attaagcatt tcatatctaa acaagttgac cggtttctgt ttcgttatgg
 39481 acattatatt gaagattttc caaggcaatg tattttctatt ggtacaacta ataaagtga tttcttaaga
 39551 tgctgaactg gtggaaagc tttttggcca atgactgtaa atccagagag agttgaagtg aactggtcta
 39621 aactaaccaa agaaagatgc gaccaaatct gggcagaagc taatatctat tatgaacaa gagaagagtt
 39691 gttccttaac cctgaactag aagaagaat cgttcaatc caaagtaaac atactgagga atctccat
 39761 acaggtatta tgaatgaata tcttaacacg tcttaaccaa gcaattggga agacttaact atctttgaa
 39831 gaagacgatt ttatcaaggt gatgttgata tgttaccaac aggaatgta gattacattg aaagagacaa
 39901 ggtctgtgog ctggaagtgt ttgttgaatg ttttggtaaa gataagggag atagtagagg atctatgaa
 39971 attagaagaa tttctaacgt ctttaagacaa tttagacaatt ggtctgtata tgaaggcaat aaaagtggga
 40041 aaattcgatt tggaaaaag tatggtgtac agatagcgta tgtaagagat gaaagtttag aggtatttat
 40111 acaagaataa ttgaataaat atacattttt agatgttgta tcaaatgctg catcattttt tgagtgtatg
 40181 aacacggttg tgtaaaaagt aatcgtaggt gttgtatcat ttttgggtgat gcaacattga tgcacaaat
 40251 gatacaacac ctctttccct tctcgttgta aggttcaacc ctgtttgttt ccaatgttgc atcaaatcca
 40321 ctataaagtg taaaaagtag tgttagggag taaaaggggt taggggtaac cctctaacag ctatttttaa
 40391 aagtttgcca agaattgat caacatcgga acacaaatat aatttttgta tacaaggtga ataaatgaaa
 40461 gaatcgacat tagaaaaata ttttagtgaa gagataacaa agttaaatgg attatgttta aaatgggtcg
 40531 cacttggaac aagaggtgta ccagatagaa ttattattat gccagaagga aaaaatattt ttgtagaat
 40601 gaagcaagaa aagggaaggt tacatctctt acaaaaatat gtgcatcgcc aatttgaaaa cagagatcat
 40671 acagtgtatg tgttatggaa taaagaacaa gtaaatactt ttataagaat ggtaggtgga acatttggcg
 40741 attgatttca aaccacatag ctatcaaaag catgcaatag ataaagtgat tgataatgag aaatcaggtt
 40811 tgtttttaga tatggggcta gggaaaaacag tatcaacact tacagcattt agtgaattgc agttgttaga
 40881 cactaaaaaa atgttagtca tagcacctaa acaagttgct aaagatacat gggttgatga agttgataag
 40951 tggaaaccatt taaatcatct gaaagtgtct ttagtcttag gaacacctaa agaaagaat gatgcattaa
 41021 acacagaggg tgatatctat gtaaccaata aagaaaatac taaatggtta tgtgatcaat ataaaaaga
 41091 atggccattt gacatggttg taattgatga actgtctaca tttaaaagtc ctaagagcca aaggttttaa
 41161 tctattaaaa agaaattacc actcatttat agatttatag gattaaacag aacacctagt ccaaatagtt
 41231 tacaggattt atgggctcaa gtttatttga tagacagagg cgaagagact gactcttcat tcagtcgta
 41301 tcaagaaggg tactttaaac caacacatca agtttagcga catgttttta actgggagct aagagacgga
 41371 tctgaagaaa agatatatga acgaatagaa gatatatgtt taagcatgaa agcgaagat tatctggata
 41441 tgccctgacag agttgatatt aaacaaacag tagtcttatt tgaaaaagaa agaaaagtat atgaagaatt
 41511 agaaaaaaac tatatttttag aatcggaaga gttgtagctc agaatggggc agaatggggc atcttaagt
 41581 caaaaactac ttcaactatc taacggtgca gtttatacag atgatgaaga tgtaagactt atacatgata
 41651 agaagttaga taagttagag gaaattatag aggaagctca agggcaacca atattattgt tttataact
 41721 caaacatgat aaagaaagaa tacttcaag gtttaaggaa gcaaccacat tagagagatt aaactataaa
 41791 gaacgttgga atagtggaga cattaaagct cttatagcac atccagcaag tgcagggcat ggtatnaact
 41861 tacaacaagg tgggcacatt attgtttgtt ttggacttact atgttctatg gaattatacc aacaagcaaa
 41931 tgcaagatta tatagacaa gacaaaaatc tacgactatt attcatcaca tcatgaccga taacacataa

183

```
42001 gatcaaagag tatataaagc tttaaaaaat aaagaactaa cgcaagaaga attgatgaaa gctatttaaag
42071 caagaatagc taagcataag taatggagggt ataagatggg aaaggcggtca tatgatatta agccagggaac
42141 atttaaatat attgaatcag aaatatataa tttaaatgag aacaagaaaag agataaataag attgagaatg
42211 gagatactta acccaacgaa agaactagac accaacattg tgtatggacc gttacaaaaa ggagagccag
42281 ttagaacaac tgagttaatg gcgacaagggt tattgactaa taagatggtta cgtaacttag aagagatggt
42351 tgagcagtt gaaagtgagt acttaaaagt acctgaagat cataagaaaag taataagggt aaagtattgg
42421 aataaagata agaagctaaa gatagaacaa ataggggatg cttgtcacat gcacgcgaat acagttacta
42491 caatacgaaa gaactttgtt aaagcgatag cgtatcatgc aggtatcaaa taacattgtg caaagattgt
42561 gcaaaaggcc tacaatatctg tagtaatatg atagtatcgg aaagatgtat aaagttatct gaaagttata
42631 cgacataaat acatgaggca catcgctaag cggtgtgtct tttgttatgc aatcaaagag gtgtaagaga
42701 tgaccaagca taataacatt tataagcatg gtcgtaagtc atatcaatac gattggttct atcattcaaa
42771 agcatggaag aagttaagag agatagcatt agatagagat aattatcttt gtcaaatgtg ttacgcgaa
42841 gatattataa cagatgcaaa gattgtgcat cacattattt atgttgatga agattttaac aaagctttag
42911 acttagataa tctaattgtca gtttgttata gctgtcataa caaaattcat gcaaatgata atgacaaaag
42981 taatcttaag aaaattagag ttctaaaaat ttaataaaaa aaattattta aataaaaattt tatgcccccc
43051 tgcccatcgg cttaaaatgt tttttcgcgg ggtaccggag aggcc
```

Table 8

Bacteriophage 3A ORFs list

SID	LAN	PRA	POS	a.a.	RBS sequence	STA	STO
100379	3AORF001	1	8515..13488	1657	acagggtacggatttaagaaaacttt	ttg	taa
100380	3AORF002	2	37667..40114	815	tttaaaataatgaaaggagccgaac	atg	taa
100381	3AORF003	1	32188..34149	653	ttaaagaaattgaggtgtcaagaat	ttg	tag
100382	3AORF004	3	17457..19370	637	gctattttattagaaaggaggtgc	att	taa
100383	3AORF005	1	334..2034	566	agaaaaagatagttcaagaagaag	gtg	taa
100384	3AORF006	1	15571..17154	527	ctttttatttataggtagggtgattta	atg	taa
100385	3AORF007	2	19337..20836	499	atgatagtaaaacaagttcagggcc	atg	taa
100386	3AORF008	3	22176..23630	484	aatgatttagggtaggtgttgacca	atg	tga
100387	3AORF009	1	40726..42093	455	gtaaataacttttataagaatggtag	gtg	taa
100388	3AORF010	3	13491..14738	415	gaggcggaactaacgctacagtaaaa	att	taa
100389	3AORF011	2	2039..3277	412	attaaagacataatgcgttaaggag	gtg	taa
100390	3AORF012	2	4001..5209	402	aaaaaagagaaaaaattaaacgcga	atg	taa
100391	3AORF013	1	30379..31545	388	attttatgaatgcgagaataaatgc	atg	taa
100392	3AORF014	2	14738..15562	274	atttatatgggaggtttgactaatta	atg	tag
100393	3AORF015	3	3249..4034	261	cttgaattaagaaaatctttgaaag	gtg	tag
100394	3AORF016	-2	25587..26273	228	aagaagctaaagaaaaataaaaaat	atg	tga
100395	3AORF017	3	6729..7370	213	ttaatttttaaggaggaaataagca	atg	taa
100396	3AORF018	3	24540..25154	204	aataaaaataaaaagtaggtgataag	atg	taa
100397	3AORF019	2	31565..32128	187	ctataaaaattaaaaaggacggtat	ata	taa
100398	3AORF020	3	36150..36713	187	gcagtaggaattatgacgggtcaag	ttg	taa
100399	3AORF021	2	24011..24535	174	gtaataaaatttataaagaaaggaa	atg	tga
100400	3AORF022	-2	12423..12938	171	taaagtcaccagtagacaatgtaggt	att	tga
100401	3AORF023	1	7462..7917	151	aaaataaatcaaaggagaataattt	atg	taa
100402	3AORF024	1	26731..27174	147	actaaataaaaaataaggaggacact	atg	tga
100403	3AORF025	1	42106..42543	145	taagcataagtaattggaggtataag	atg	taa
100404	3AORF026	2	35255..35671	138	aagcaactaactttttttaaggag	ata	taa
100405	3AORF027	2	5888..6298	136	attttggctataatacagctgggttt	atc	taa
100406	3AORF028	-3	27845..28255	136	ctttttaagatgtttatgatccttt	ctg	taa
100407	3AORF029	3	34344..34748	134	tttaagggttttagatttagaggtgga	atg	taa
100408	3AORF030	2	6299..6694	131	tataaaaaaggaggttgccagataa	atg	tag
100409	3AORF031	1	20833..21225	130	ttacaaaaattataggagtgagaaa	ata	taa
100410	3AORF032	-2	39984..40361	125	aaatagctgttagagggttaccct	ata	tag
100411	3AORF033	1	7957..8325	122	gaatatctgcgtcttttttatttga	ata	taa
100412	3AORF034	-2	28506..28871	121	gttatcaacctaaaggaggtgataac	atg	tag
100413	3AORF035	-2	10671..11036	121	tcctagcttccctaacagcaccgcca	ata	tga
100414	3AORF036	2	30020..30382	120	accaatttttaaggaggaggttaatca	atg	tga
100415	3AORF037	2	21818..22165	115	aagtgttaagtaatagtttaagagcca	gtg	tag
100416	3AORF038	-2	42003..42347	114	gtactcactttcaactgcttcaacc	atc	tga
100417	3AORF039	2	21386..21727	113	tccagaaaaatctagagtcataaggtt	ata	taa
100418	3AORF040	-3	29654..29995	113	ttgattaactcctccttaaaattgg	ttg	taa
100419	3AORF041	-1	4333..4671	112	tactaaatctacatctgatccatga	att	tga
100420	3AORF042	3	5568..5900	110	taaaaaagtggttaggtgatttttaa	atg	tga
100421	3AORF043	1	25690..26019	109	taccaaattaatatagctcttcgcat	ata	tag
100422	3AORF044	3	29676..30005	109	gtcttaataattatataaggaggtt	att	taa
100423	3AORF045	3	30..353	107	cgctagcaacgcggataaatttttc	atg	taa
100424	3AORF046	3	27894..28214	106	aagatattgaaaagctaatttcccc	ata	tga
100425	3AORF047	-2	11907..12227	106	ttcgccgcaaaatgattagcattt	ctg	tga
100426	3AORF048	-3	40343..40663	106	ccataacacatacactgtatgatct	ctg	taa
100427	3AORF049	-3	6749..7069	106	tggttaaacatctttagattctcca	ata	taa
100428	3AORF050	1	42700..43014	104	ttatgcaatcaaagaggtgtaagag	atg	taa
100429	3AORF051	-2	13077..13388	103	ttgtacgtaatcccacacatcgccg	att	tga
100430	3AORF052	-3	3722..4024	100	gcatttcatttctcctaataactc	att	tga
100431	3AORF053	3	17145..17444	99	tccagacaatggatagggaggtgt	att	tag
100432	3AORF054	-1	19915..20211	98	ataattttatagcttgcgaacataa	ata	tga
100433	3AORF055	-1	42436..42729	97	aatcgtattgatatgacttacgacc	atg	tag
100434	3AORF056	3	40455..40745	96	taaaattttgtatacaaggtgaataa	atg	tga
100435	3AORF057	-1	38665..38952	95	atcatcacgctcttgccattgacgt	att	taa
100436	3AORF058	-1	21265..21549	94	gaaattttctatctaacttgcataa	att	tga
100437	3AORF059	-2	10278..10562	94	tttagccgcgcttccaactgcacgt	att	tag
100438	3AORF060	1	5278..5556	92	atatcagccgaataggggtgatgaa	atg	tag
100439	3AORF061	1	35668..35946	92	tttggaagaagggaggttgattaa	ata	taa
100440	3AORF062	2	35912..36187	91	gttaaatctggaatggaattaaaca	ata	taa

100441	3AORF063	3	36720..36995	91	cggaagtagcggagctgtaaaagacat	att	tga
100442	3AORF064	-2	35694..35969	91	ccgttatacgcgctagcactaataa	ctg	taa
100443	3AORF065	-2	32697..32972	91	aaccgttttcttttgtaaatagggt	ata	taa
100444	3AORF066	3	29157..29429	90	caaaccttaacatttatctaaagga	gtg	tag
100445	3AORF067	-2	26661..26930	89	atacttttttagcgggaatcggtga	ttg	taa
100446	3AORF068	-2	9624..9893	89	ttttaatgcacatccccatgtattga	ata	tga
100447	3AORF069	-3	13847..14110	87	tgcatcttccctcctgattcgtgtga	atc	tga
100448	3AORF070	1	34993..35250	85	tttacgtccaaagagcttttgactt	gtg	taa
100449	3AORF071	2	34745..35002	85	aaatgttcaagaaatggagtgaaagc	ata	tga
100450	3AORF072	-1	27379..27636	85	tttgcgttctctcctttaagttggt	ttg	taa
100451	3AORF073	2	37367..37615	82	tggtaatagctatttatcatttttga	att	taa
100452	3AORF074	-2	23466..23714	82	cgtttgcgttttttaaaatttaatat	att	taa
100453	3AORF075	-3	2471..2719	82	agtactgtttgaaatcttctaaccac	ttg	tga
100454	3AORF076	1	26047..26292	81	aagtagcgttttcttggcggggagggt	gtg	tag
100455	3AORF077	2	28292..28537	81	aacatcttaaaaggagggaataacaa	atg	tag
100456	3AORF078	-1	5836..6075	79	tttgtataaggcttagatttagtc	att	taa
100457	3AORF079	-2	5460..5699	79	attcagtcgctttttaaaatttctct	atc	taa
100458	3AORF080	-2	31350..31586	78	cctgtaatcacttttagtttttatta	ata	taa
100459	3AORF081	-3	8252..8488	78	aagttttcttaaatccgacacctga	atg	tga
100460	3AORF082	-1	35905..36138	77	atatttatagacaacttgaccgcgtc	ata	taa
100461	3AORF083	-1	34039..34272	77	atagttcacctggattattaaataa	ata	tga
100462	3AORF084	-1	12007..12240	77	acatttttttctatttcgccgcctaa	atg	taa
100463	3AORF085	-2	32367..32597	76	cttacaagggtatagagaaataacga	att	taa
100464	3AORF086	-2	30618..30848	76	atataatctaaagtgaggattatct	ata	taa
100465	3AORF087	-3	24746..24973	75	ataggtttttaagttcacctcttca	atg	tga
100466	3AORF088	-3	12980..13204	74	tctttcttttctgtaaccacatgga	att	tag
100467	3AORF089	3	4290..4508	72	acaggagaagccttatcaatctttaa	atg	taa
100468	3AORF090	3	28926..29141	71	ttatacacgaaaggagcataaaca	atg	taa
100469	3AORF091	-2	13587..13802	71	ctgtctctgctaattgcttagataa	atg	tag
100470	3AORF092	2	26471..26683	70	aaacgaacacaaaggagggtttca	atg	taa
100471	3AORF093	-1	2524..2736	70	tccaccgttttcttcatagtagtct	ttg	tga
100472	3AORF094	-3	25334..25546	70	tggcgcttttaataataaaagcgtct	att	tga
100473	3AORF095	3	8316..8525	69	aagagatgggaaagacagaagaaca	atc	tag
100474	3AORF096	2	36992..37198	68	aaacaagttcaaggaggctatgaggga	atg	tga
100475	3AORF097	-1	32593..32799	68	aaagcttaataacctctgtcgtttat	atg	taa
100476	3AORF098	-1	15346..15552	68	aatccattaaatcacctacatataa	ata	tag
100477	3AORF099	1	7225..7428	67	actggtgactggatgaacagaaaaag	ttg	tag
100478	3AORF100	-2	22620..22823	67	cgacttcacgacccgcatgtcttaa	ata	taa
100479	3AORF101	-1	40060..40260	66	aaccttacagcgaaggggaagagga	gtg	taa
100480	3AORF102	-1	35035..35235	66	ttctatctctttaaataaaagttag	ttg	taa
100481	3AORF103	-2	1149..1349	66	atttttttggagtggtgggtaatca	ata	taa
100482	3AORF104	1	27661..27858	65	aaacaacttaaaaggagggaacgacaa	atg	tga
100483	3AORF105	-2	9420..9617	65	gcctaagtcaaccgcttgattagac	atg	tga
100484	3AORF106	-2	23244..23438	64	caccagtaattcttgaattagttga	ata	taa
100485	3AORF107	2	11966..12157	63	tctaaaaaagatgctgtagtagacg	ttg	taa
100486	3AORF108	-3	35054..35245	63	ttttcatcatttctatctccttaaa	ata	tag
100487	3AORF109	-3	16010..16201	63	gttcttaattccaagtactgacag	ttg	taa
100488	3AORF110	-1	6184..6372	62	attttcagtgactttataatagtat	att	taa
100489	3AORF111	-2	16500..16688	62	gtagtcaacaattgctttgtattga	ttg	tga
100490	3AORF112	-2	8502..8690	62	cttaattctcgctgatacttttcc	att	taa
100491	3AORF113	1	34162..34347	61	tatgaaggattaggagtggtgattgc	atg	tga
100492	3AORF114	2	12356..12541	61	ggatatcacactaaggctatagcta	ata	taa
100493	3AORF115	-2	7635..7820	61	tgaagttccctcagctacaccgtga	att	tga
100494	3AORF116	-1	26434..26613	59	tttagcttctgaaagttgtaaaatct	ctg	tga
100495	3AORF117	-3	17804..17983	59	atagccattatttctagcttgrgtc	atg	tga
100496	3AORF118	2	27899..28075	58	attgaaaagctaatttccccataag	att	taa
100497	3AORF119	-1	39268..39444	58	acgaaacccggtcaacttggttagat	atg	tga
100498	3AORF120	-2	37152..37328	58	tagctattaccatgaaacttcagct	ctg	taa
100499	3AORF121	-2	18900..19076	58	aaggtagctctctcccatttaccact	att	taa
100500	3AORF122	-1	21550..21723	57	taagcatggtaatcacctccttttaa	atg	taa
100501	3AORF123	-3	33062..33235	57	aaacgttggtctttaaataagatctc	ttg	tag
100502	3AORF124	2	21212..21382	56	aaatagaagagggttaaaggagaga	ctg	tag
100503	3AORF125	-1	22051..22221	56	aaatcaggattgaactgcttcctta	atg	tga
100504	3AORF126	-2	7821..7991	56	tgtttttctctgttttacgggtcttta	att	tga
100505	3AORF127	-3	34712..34882	56	ttgcattacatattgcgaatgctag	ttg	taa
100506	3AORF128	-3	24056..24226	56	tttttaaaatcaaagcgtctttgttt	ata	-taa
100507	3AORF129	-3	4940..5110	56	cataccatgcagttaatacaacaa	ata	tga
100508	3AORF130	3	27171..27338	55	cagaattaaactatcgattgttcga	atg	taa
100509	3AORF131	-1	40387..40554	55	ccttctggcataataataattcttat	ctg	taa
100510	3AORF132	-2	1860..2027	55	gcgataacattcacctccttaacgc	att	tga
100511	3AORF133	-3	42317..42484	55	acaaagttctttcgtattgtagtaa	ctg	tag
100512	3AORF134	2	12671..12835	54	tcatacaaatctttaaaagggtgga	ctg	tag

100513	3AORF135	-1	39484..39648	54	ataatagtatttagcttctgcccag	att	taa
100514	3AORF136	1	29710..29871	53	accttacaacaaaaatactatcac	att	taa
100515	3AORF137	1	37186..37347	53	ggcagttgtttgaaaataaaggga	gtg	taa
100516	3AORF138	2	20996..21157	53	aatggggaaatagtttttaacgaag	att	taa
100517	3AORF139	3	15114..15275	53	tcaactgaaattgaagtaagttaa	atg	taa
100518	3AORF140	3	29442..29603	53	aaaatgggtattaggaggattatcaa	atg	taa
100519	3AORF141	-1	39883..40044	53	tacaccataatcttttccaaatcga	att	taa
100520	3AORF142	-1	20416..20577	53	accacctggaaaagtcctcataaaaa	att	tga
100521	3AORF143	-1	1942..2103	53	ataaagcttagaagttgactgatca	atc	taa
100522	3AORF144	-3	39380..39541	53	ttccaccagtttcatctcttaagaa	atc	taa
100523	3AORF145	3	20388..20546	52	tctgagtggtcagaattagctatta	atg	taa
100524	3AORF146	-2	2358..2516	52	aacatgtccatattatgaacaatca	att	tga
100525	3AORF147	-3	5606..5764	52	gtgatttgtttgtggtagatattca	att	tga
100526	3AORF148	2	34145..34300	51	tttacttctccgtttttatatgaagg	att	taa
100527	3AORF149	-1	7918..8073	51	cattctcttgatttactaattcttaa	ata	taa
100528	3AORF150	-2	11745..11900	51	ttcatccttatgtctttgatcagca	ata	taa
100529	3AORF151	-3	7097..7252	51	ttaccttcatgatacccgatataca	ata	tga
100530	3AORF152	1	21652..21804	50	ctaaaaatatttagagacatcaagat	gtg	taa
100531	3AORF153	2	5381..5533	50	tcggctaagctcgaattactattaa	gtg	tga
100532	3AORF154	-1	39670..39822	50	ttgataaaatcgtcttctttcaag	ata	taa
100533	3AORF155	-1	38233..38385	50	ataggctctacaaaatgcaccaaca	att	tag
100534	3AORF156	-1	33040..33192	50	tatctgaaatataatgctttgttaa	att	tag
100535	3AORF157	-2	10119..10271	50	cttcaatgatttgcctatagctatta	att	tga
100536	3AORF158	-3	36074..36226	50	atccgtcttatgatttcgttctggc	att	taa
100537	3AORF159	-3	18338..18490	50	taaatagtttctattattctggatta	ctg	taa
100538	3AORF160	3	39399..39548	49	gtttggttaagtgaatggcagaac	ttg	taa
100539	3AORF161	-2	8976..9125	49	ttgtacttttagtttttgaacttga	ttg	tga
100540	3AORF162	-3	31199..31348	49	tctgtaatatcttcaggtttataac	ctg	tga
100541	3AORF163	-3	14459..14608	49	attactctgagaagaacagtttga	atc	tga
100542	3AORF164	3	25182..25328	48	ttttttcttagctttttctgataaaa	gtg	tag
100543	3AORF165	3	28353..28499	48	aatcttgtctctatgacacggaaag	att	taa
100544	3AORF166	-1	8899..9045	48	gtactgcgtcacttgcctcttttag	ttg	taa
100545	3AORF167	-2	411..557	48	taatacaagttgacgttttagatcct	ttg	tga
100546	3AORF168	-3	25973..26119	48	gctgagtagtctcaatgtggaagatta	atg	tag
100547	3AORF169	-3	25151..25297	48	aaaaaaaaacgcctacaagtgtgacg	ttg	tag
100548	3AORF170	-3	24995..25141	48	taagaaaaaagatttagtattcattc	att	tag
100549	3AORF171	1	23437..23580	47	aaaggttaataacgtaaggacggct	att	tag
100550	3AORF172	2	32414..32557	47	ctattttgacctgtctgtaaaaaagt	atg	taa
100551	3AORF173	-1	38005..38148	47	ataagttgtatcatcgaagtaattcc	atg	taa
100552	3AORF174	-1	4123..4266	47	atttaaaagttagataagctttctcct	gtg	tga
100553	3AORF175	-1	3124..3267	47	ttcattttgaaaatacttagctttca	ttg	tag
100554	3AORF176	-1	580..723	47	cattttctccatcttgtgatacagc	ata	taa
100555	3AORF177	-2	39819..39962	47	ttgaaaatcttttctaatttccatag	atc	tag
100556	3AORF178	-2	38466..38609	47	ttagcgtcttcatcttgagcaccat	ata	tag
100557	3AORF179	-2	33927..34070	47	ttttgcccatacaacaggcttattc	atg	tga
100558	3AORF180	-2	33555..33698	47	cgtcttttcgggattttacagtatta	att	tga
100559	3AORF181	-2	29538..29681	47	atagtattttttgttgaaggtcat	att	tga
100560	3AORF182	-3	17099..17242	47	aatatcactactgcctgcataaggt	atc	tag
100561	3AORF183	2	23750..23890	46	ttaaaaaaacaaacgttttttagtat	ata	taa
100562	3AORF184	-1	31648..31788	46	tggaagtttcagatttgcaggaact	ttg	tga
100563	3AORF185	-1	30565..30705	46	attttgtttcaaataaagctattac	atc	tag
100564	3AORF186	-1	16951..17091	46	gagaattcaaaagtactagtgtataa	atg	tga
100565	3AORF187	-1	7153..7293	46	tatccaaacgaataacttttttgaga	att	taa
100566	3AORF188	-1	1237..1377	46	ccagctcttctaaagaaacaatttc	att	taa
100567	3AORF189	-2	33309..33449	46	catttgagaagccgatgcttcatat	atc	tga
100568	3AORF190	-2	7197..7337	46	gtaacgaacttgcagaatcctctga	atg	taa
100569	3AORF191	-3	41459..41599	46	tcactctgtataaaactgcacggttag	ata	tag
100570	3AORF192	3	4863..5000	45	gatgctattattaaacgcttttagcag	att	tag
100571	3AORF193	3	25965..26102	45	tatacgatactagtttagactcttt	ata	tga
100572	3AORF194	-1	37069..37206	45	ctagtaagaataataatcttagtat	ttg	tga
100573	3AORF195	-1	11749..11886	45	tttgatcagcaatagctaataattt	atc	tga
100574	3AORF196	-2	40764..40901	45	atcttttagcaacttgttttaggtgct	atg	tga
100575	3AORF197	-2	31989..32126	45	ggctaaaaacccaccatttgactt	ata	tga
100576	3AORF198	-3	36431..36568	45	tttatttatgacataactaccattc	ata	tga
100577	3AORF199	-3	33515..33652	45	ttccaaaaataaactatgttaggat	ttg	tga
100578	3AORF200	-3	21233..21370	45	ataagattataacctatgactctag	att	tga
100579	3AORF201	1	23293..23427	44	aagcctatcgggtggtgatatctta	gag	taa
100580	3AORF202	-1	39088..39222	44	atagtcaaaattacatcctggctcc	att	taa
100581	3AORF203	-1	16309..16443	44	tttgcttgccgtctaaaaatcaactt	ata	tga
100582	3AORF204	1	23845..23976	43	atgtttattatcaatcaaaatataa	att	taa
100583	3AORF205	1	29500..29631	43	gtgttgtgtcttcacggtcttagcga	ttg	taa
100584	3AORF206	2	16667..16798	43	gaaaaatcaacagttcttaatttaa	ttg	tag

100585	3AORF207	-1	35386..35517	43	tcgagatttatagactccttcttga	atc	taa
100586	3AORF208	-1	30013..30144	43	cagttgagctgtttcatcttttggc	att	taa
100587	3AORF209	-1	28366..28497	43	taattcctggctctctagtgggttt	ata	tga
100588	3AORF210	-1	15739..15870	43	catcaagcttatttgattccactga	gtg	tag
100589	3AORF211	-1	7693..7824	43	taactgaagttccctcagctacacc	gtg	tga
100590	3AORF212	-2	4314..4445	43	ggttctgaaacaatttctttagaaa	gtg	tag
100591	3AORF213	-2	4011..4142	43	tgtttgatgtcttccatcatcaatat	ttg	taa
100592	3AORF214	-2	1722..1853	43	ctgtctagtttcaactgaacatta	ttg	taa
100593	3AORF215	-3	16616..16747	43	tcttcatttgtttgctgatttagcat	atc	tag
100594	3AORF216	-3	15833..15964	43	gtcattttgaccgaagttttttgat	ttg	taa
100595	3AORF217	3	6363..6491	42	gatgcagagctccaaacatatataa	att	taa
100596	3AORF218	-1	32146..32274	42	aataagctataattaagatttcgaa	atc	taa
100597	3AORF219	-1	29800..29928	42	ctagggtcatcactttgttcgttta	atc	taa
100598	3AORF220	-1	18409..18537	42	gcattaacctgatagccttcttctc	ctg	tag
100599	3AORF221	-1	13234..13362	42	ttttatcgcttaaccaagatgcacc	atc	tag
100600	3AORF222	-1	12313..12441	42	cccaagctttatctgagggcatctga	ata	tga
100601	3AORF223	-1	4915..5043	42	tccatcatagtttaattccaaaaataa	ttg	taa
100602	3AORF224	-1	2125..2253	42	attaactactttataactctcatcac	att	taa
100603	3AORF225	-2	26298..26426	42	tcgtttgtaacaacttgatttaaga	ata	taa
100604	3AORF226	-2	17184..17312	42	cgcttatttttaattatctaatttt	att	tag
100605	3AORF227	-2	1425..1553	42	atcttcttccattctctataggggt	att	taa
100606	3AORF228	-3	31055..31183	42	cattttttgatgtcaggcagtttat	ata	taa
100607	3AORF229	-3	22592..22720	42	gttataacatgacggctacaaagc	ata	taa
100608	3AORF230	-1	27883..28008	41	gaaggcaggggtcgtttcttggatta	ttg	tag
100609	3AORF231	-2	29988..30113	41	gcttcttttaactttctcttgacaaa	ttg	taa
100610	3AORF232	-2	22485..22610	41	tatctgggaaatttaacttaataaaa	ata	tag
100611	3AORF233	-2	9264..9389	41	aagtttgccgaatgactttgagct	atc	tga
100612	3AORF234	-3	23033..23158	41	acctaatcagataaagcgataattt	ata	tga
100613	3AORF235	1	25558..25680	40	aacactgctgaatatagacgtctttt	ata	tag
100614	3AORF236	1	34420..34542	40	acattgagagaagtttcagaaaaat	atc	taa
100615	3AORF237	3	38442..38564	40	gaagaagctatagaaacttttattc	ctg	taa
100616	3AORF238	-1	33628..33750	40	caatcattagaaaacttttttcat	ata	taa
100617	3AORF239	-1	29248..29370	40	tcttctaatttagaaaattataatca	atg	tag
100618	3AORF240	-2	18156..18278	40	gtctctcaattctgtatagaattttt	att	taa
100619	3AORF241	-2	8088..8210	40	tttcaaggctttttgtataagtttta	gtg	tga
100620	3AORF242	-3	39149..39271	40	ttagcaagcagatttacctacacc	ttg	taa
100621	3AORF243	-3	23558..23680	40	aaaatttaactgtttatcaattttaa	ata	taa
100622	3AORF244	-3	1697..1819	40	catcttcaataaaggattattattaa	ata	tga
100623	3AORF245	1	19015..19134	39	agttatgcaaggaatatgatgactt	ttg	tag
100624	3AORF246	1	22504..22623	39	gctaactctaaacactttcacatcgt	ttg	taa
100625	3AORF247	-1	40567..40686	39	aaagtatttacttgttctttattcc	ata	taa
100626	3AORF248	-1	23956..24075	39	tttagattcatgaaacgaagttagca	ata	taa
100627	3AORF249	-1	11113..11232	39	cacctttcccaacacttttaccagt	atc	tga
100628	3AORF250	-1	8719..8838	39	ttttattagcttctactagcttttaa	ata	taa
100629	3AORF251	-2	16899..17018	39	aactcgtctgttaagcgcttgttga	att	tga
100630	3AORF252	-3	37025..37144	39	acaactgccctaatttaataactgc	att	tga
100631	3AORF253	-3	29138..29257	39	tctacatactccaacaattgtatgg	att	taa
100632	3AORF254	-3	15476..15595	39	caaatcaattcattaaaatccatta	ctg	taa
100633	3AORF255	1	13552..13668	38	ttaatagacaaagttaaaatcgtggt	ttg	tag
100634	3AORF256	2	12545..12661	38	aaaagtgcgaagggtcgtgtaacgg	ata	taa
100635	3AORF257	2	41870..41986	38	gggcatggattaaacttacaacaag	gtg	tga
100636	3AORF258	3	10827..10943	38	tcaaaacttttgaaaaacgggttagg	att	taa
100637	3AORF259	-1	34570..34686	38	gtgacatcgaaccagtagcgatcac	gtg	tga
100638	3AORF260	-1	32389..32505	38	aagcaggttaagccaatacgcattga	att	tag
100639	3AORF261	-1	23830..23946	38	ccttttttaacttttaataaaattaa	ata	tga
100640	3AORF262	-1	8158..8274	38	ccatctcttctgggttcagtttctga	atc	taa
100641	3AORF263	-2	14001..14117	38	ttatacctgcatttctctctgattc	gtg	tga
100642	3AORF264	-2	294..410	38	tttgcttgtttttattttcccttga	gtg	taa
100643	3AORF265	-3	42683..42799	38	tgacaaagataattatctctatcta	atg	tga
100644	3AORF266	-3	31979..32095	38	aatcctcatcatcagtgcttaattc	atc	taa
100645	3AORF267	-3	26306..26422	38	ttgtaacaacttgatttaagaatac	atc	tga
100646	3AORF268	-3	16490..16606	38	tacatacaaggcttagcttttttat	ttg	tag
100647	3AORF269	-3	9872..9988	38	tgagaccctctaacctgagtttag	ata	tag
100648	3AORF270	1	21829..21942	37	atagttaagagtcagtgcttcggca	ctg	tag
100649	3AORF271	2	29468..29581	37	tgagcgacacatataaaagctacct	att	taa
100650	3AORF272	3	2955..3068	37	gaggttaaacagattttacttgcagc	ata	taa
100651	3AORF273	3	5010..5123	37	tttggaaccagtagtatttacag	atg	taa
100652	3AORF274	3	19956..20069	37	tcaagtatagatgaattaaagcaac	ttg	tga
100653	3AORF275	3	39882..39995	37	gatattgttaccacaggaattgtag	att	taa
100654	3AORF276	-1	27211..27324	37	attaagtgcgcttattttaattagat	att	tga
100655	3AORF277	-1	13516..13629	37	cgaccgtcattaaagttaagtcac	ctg	tga
100656	3AORF278	-1	11893..12006	37	ttttatatacacgaccactggataa	atc	taa

100657	3AORF279	-2	17535..17648	37	tttgtaaagatttgtttactgctgc	ttg	taa
100658	3AORF280	-2	6474..6587	37	tcaaaataagcatctaactgactag	atg	taa
100659	3AORF281	-2	759..872	37	ttttgatatcggttgcgtcataatgg	att	tga
100660	3AORF282	-3	36608..36721	37	cccaaaacctcttctgactcgatcta	ata	tga
100661	3AORF283	-3	14960..15073	37	tttcagtttgaagaacctcttttaa	att	taa
100662	3AORF284	1	18859..18969	36	atgttaacagagccaggtctttact	att	taa
100663	3AORF285	2	8237..8347	36	aaaacttatacaaaagccttgaaaag	ata	taa
100664	3AORF286	3	5157..5267	36	tatgatcagcaacgtacattagaca	gtg	tag
100665	3AORF287	3	38610..38720	36	tttgatttagtacgcatacacttat	atg	taa
100666	3AORF288	-1	36454..36564	36	tttatgacataactaccattcatac	ata	tga
100667	3AORF289	-1	30217..30327	36	aacaattttttcataatgctcttct	ttg	taa
100668	3AORF290	-1	16678..16788	36	gcttttttgcaatttctaacagctt	atc	tga
100669	3AORF291	-2	14310..14420	36	gtctagttaaaggataaacatctc	ctg	tga
100670	3AORF292	-2	11457..11567	36	ttctttcaattctttgattttctga	ttg	tga
100671	3AORF293	-3	29462..29572	36	ttcataaaagtattccttataaaat	atg	tag
100672	3AORF294	-3	22388..22498	36	accattccaatttttgccaaacgat	gtg	tag
100673	3AORF295	-3	18629..18739	36	aaaaggaacgcctcttgagtgaaat	att	tag
100674	3AORF296	-3	6332..6442	36	tatcagacatgaagtcrgaaggtaa	atc	taa
100675	3AORF297	1	13984..14091	35	aaatggttgaagtcacttaaaggta	gtg	tag
100676	3AORF298	1	40174..40281	35	tatcaaatgttgcacattttttga	gtg	taa
100677	3AORF299	2	1481..1588	35	gccgctgtgctacttttgcgttag	ata	taa
100678	3AORF300	2	40451..40558	35	aataataaattttgtatacaagggtga	ata	tag
100679	3AORF301	3	25479..25586	35	accactagttaaaacttcataact	ata	taa
100680	3AORF302	3	32106..32213	35	gaagatgatttcgatgaattagaca	ctg	tga
100681	3AORF303	3	36024..36131	35	gacacagagggattattaaaagaga	ttg	tag
100682	3AORF304	-1	37762..37869	35	accgacaaatccgccaaactccttt	ata	tga
100683	3AORF305	-1	24088..24195	35	tttatcttttaacaaactcaactga	ata	tga
100684	3AORF306	-1	19507..19614	35	atcattaggttaattgaaattttaaa	ata	tga
100685	3AORF307	-1	16081..16188	35	atgtactgacagttgcagatacagt	atc	tag
100686	3AORF308	-1	11398..11505	35	tttcttttagtcttctagttaaattgtt	ttg	taa
100687	3AORF309	-2	33003..33110	35	aaacagacctcttaccgcttcatca	ctg	taa
100688	3AORF310	-2	24894..25001	35	gtaatcgaatcgctaccagctga	att	taa
100689	3AORF311	-2	22005..22112	35	ttcgtaggtgtcattactttcttaa	ttg	tag
100690	3AORF312	-2	21711..21818	35	aaaataaaaagccagtgccgaagca	ctg	tag
100691	3AORF313	-2	17901..18008	35	cattaggtcttagacgacttagcat	ata	taa
100692	3AORF314	-2	16710..16817	35	taattcagctcttaggagatcatctt	att	tag
100693	3AORF315	-2	15990..16097	35	acatatctccgtatcatttgggttaa	att	tag
100694	3AORF316	-2	2862..2969	35	aattctctctcatactggttgacga	ttg	tag
100695	3AORF317	-3	40217..40324	35	tccttaacactactttttaaacttt	ata	tga
100696	3AORF318	-3	37535..37642	35	tgttcggctcctttcattatttttaa	ata	taa
100697	3AORF319	-3	34421..34528	35	ttcttcactcttttatttgactctgc	ata	tga
100698	3AORF320	-3	28262..28369	35	catttgttggtaatatcttagttcg	atg	tga
100699	3AORF321	1	23989..24093	34	taaaaagggttaatatataaaatgta	ata	tga
100700	3AORF322	1	34660..34764	34	aagagaagatttgagaccatggcttt	atg	taa
100701	3AORF323	3	30105..30209	34	ctaataactgaaactatcaactgtag	att	taa
100702	3AORF324	3	30258..30362	34	ggaaaagagttccttaaaaagcag	ata	tga
100703	3AORF325	3	40236..40340	34	gttgatcatcttttgggtgatgcaac	att	tag
100704	3AORF326	-1	36964..37068	34	cgcatcaacaactgtaaacctttga	ttg	tga
100705	3AORF327	-1	35242..35346	34	atttttgcctgtgtgataatatttt	ctg	taa
100706	3AORF328	-1	21916..22020	34	ccatttaccttcttgagatggtgga	ttg	tga
100707	3AORF329	-1	18820..18924	34	ggtggcttaacttccaagaaccaac	ctg	taa
100708	3AORF330	-1	15631..15735	34	ttatgaagttttcacaaattagtaa	atc	tag
100709	3AORF331	-2	37998..38102	34	ttacgcccataagcttcatactcat	ctg	tag
100710	3AORF332	-2	7359..7463	34	tttataaaccttttaagtttttagtc	ata	taa
100711	3AORF333	-3	24584..24688	34	aaaattataaaactataaaaccat	atc	taa
100712	3AORF334	-3	24269..24373	34	tatttttaggtagataatttatttaa	atc	tga
100713	3AORF335	-3	14273..14377	34	cacttcagcaagttgatgctttgta	atc	tga
100714	3AORF336	2	7559..7660	33	gtaactttatctaatttagaagcgg	ata	tag
100715	3AORF337	2	13277..13378	33	aatataggtaaaaagcaggagaat	ttg	tag
100716	3AORF338	3	9501..9602	33	taggacgtacgatgacgatgggcgt	atc	taa
100717	3AORF339	3	27348..27449	33	atatctaattaaaataagcgcactta	att	tga
100718	3AORF340	-1	37372..37473	33	ttctatgggtttctatcttatgagaa	atg	taa
100719	3AORF341	-1	33421..33522	33	aagcraattcggacacttttccttt	ttg	taa
100720	3AORF342	-1	29047..29148	33	tttggcatctctatcactccttttag	ata	taa
100721	3AORF343	-1	7549..7650	33	atgatcgcctgagactagaattgg	att	taa
100722	3AORF344	-1	7297..7398	33	ctgctgaaactgttgagattttga	att	tga
100723	3AORF345	-2	23850..23951	33	ttaaaaccttttaacttttaataaaa	att	taa
100724	3AORF346	-2	20607..20708	33	aaagatgtacgactagatttagtta	atc	taa
100725	3AORF347	-2	14175..14276	33	atctgttggttaaaagaacgctaata	ctg	taa
100726	3AORF348	-2	6984..7085	33	cgtacactggttgacctgttaaac	atc	tag
100727	3AORF349	-2	6882..6983	33	tagaacgaccaataactgtatttag	atc	taa
100728	3AORF350	-3	40748..40849	33	aactgcaattcactaaatgctgtaa	gtg	tga

189

100729	3AORF351	-3	38345..38446	33	ggttagtagaatggttttcgrataa	atc	taa
100730	3AORF352	-3	38081..38182	33	tagttgaaggccaatacattaacct	atg	taa
100731	3AORF353	-3	35432..35533	33	tagcattctcatatgatgcagattt	ata	taa
100732	3AORF354	-3	34952..35053	33	ttatcctgatacagatatctcttag	atc	taa

Table 9

Bacteriophage 96, complete genome sequence

1	catagttata	ggcttttcag	ctatatacca	agataagatt	ratcccgccg	tctccataaa	aatatgcttg
71	gaaaccttga	tttaattggg	ttttaattcta	gcaagtgtca	aatatgtgtc	aagaaaataa	ttttctgaca
141	cgttgacctt	gctctttttt	atgttcatca	agtaagtgag	agtaggtgtc	ttaagttata	gatataattat
211	aatggcctaa	tcttttgcta	atataattcaa	taggtatacc	tttagaaagt	aggaaagatg	tatgcgtgtg
281	tcttaataaa	taaggtgtta	tgttagtacc	atttagtctt	attcgactct	tagcatgggt	aaatgacttt
351	tttaacggcat	tatgactcaa	tttaaacacac	ttattatctg	tacgttttgg	taattttgat	aattttagctt
421	taatatgttg	tatatctctt	tttgytacct	ccacaagtct	gtccgcgtta	actgtttttg	ttccacgaag
491	atgtattgta	ccctcttttt	cgtttagatc	gataggcaac	atattaatta	catcgctgta	tcttgcacca
561	gtgatagcta	ggatgaataa	aaaaatataa	ctcgattcgt	ctctagattt	aaagtattct	atcaattgca
631	agtattgttc	tatgggtgatg	aattttagagt	gttcgctctt	tgattttttt	gtaccacgaa	tatctatttg
701	atagctaggg	tctttcttta	aatagccctc	atataactga	tctctgaagc	attgtgataa	acaactgttt
771	aatttacgaa	ccgttttcatt	agtacgacct	cgaccgaatt	cgttcaaaaa	cttttgatac	tccgaacggt
841	tgatgttttt	tattaaaaaa	tcactcccgga	aatattcgtt	aaataatttt	aatgaacgct	gataccaata
911	gaattgtgtg	gaagcgacat	gtttcttatt	ttttgaatct	aaccaatcat	tgaatatttc	ttcaaaacttt
981	ttattttcat	ctaaattgtt	tccatcatcc	aaatctctaa	cgagtgtgtg	agcagcgttg	gttgcctcag
1051	ctttagtttt	gaatccctgac	tttcttttct	ttcttgattt	gaaagacgga	tgttttacgt	cgtactgcca
1121	agatgcgttg	gttttattct	tccttttttg	aatgttaaat	gacgcctatt	tacttttctt	ccctcaaaatt
1191	ggcaaaaaat	aataagggtta	ggcgagctac	ccgaaatttt	attgttgaac	aactattgtct	tcacttcttg
1261	cttttctctac	ttcttttcta	aaactatcat	atgattgatt	aggggtgtgt	aacgacattc	ctggaccacc
1331	tccagcatgt	tggtttttgt	ccggattatt	ttccatttct	tcagtgtgtc	ttttagcatt	taaatattct
1401	tcgtaactag	gttcgtttgg	gtcgcgtggg	tgtgcttgtt	gtccattatt	ggtagctgga	agattcttct
1471	gtacctgttg	cttagatgtg	ttattgtgtt	gttgattgtt	gttaattgtt	gtgtgttctt	cgttgtttac
1541	ttgattattg	ttatcgtttt	gattactatt	ttcttttttc	gcttctgctt	tacttttagt	ttcttctctt
1611	ttgtctttgt	tctcttctct	tgtttcggtt	ttcttgcttt	cccttttctt	atcgccgtcg	ttgtaccogc
1681	atgcacctaa	cactaacgca	ctagcttaata	ataaaactaa	taactctttc	atgttttaca	ctcctttatt
1751	tgctatttgt	tttaataaat	ctatgatctc	attgttttgt	tctatgattt	tgttttcatt	tttaagatgt
1821	tcgtctaaac	tctctattaa	gacgaaattt	tgatttatca	tttcgtaagt	aaacatttga	cctgtgttgt
1891	taggattaga	aaacgaacta	ctgaaacgcg	ttgaaaagct	atctataaat	tgaccaactt	tattttttaa
1961	taacatatct	ttaccgctct	cagacattgt	atttagttcg	cgcttattta	aagttttttc	tataattttg
2031	tatttctgtt	cctgatttct	ttcgatttct	tctactcaa	aagggaattt	gttattaaat	ttttcgataa
2101	tatcacgttt	ttcagaaact	gacatacgat	caaatacttg	tttttgacct	ttatttaact	ttcctcgaat
2171	tttccgggca	gtccaagact	cttttaactg	taacttatca	ttaggaaactt	gattcatctt	ttatatgact
2241	cttctttctca	tattttctta	tatttaaaaa	ctctcaacgg	ctcaaatgta	atcgaaactt	cgccatagtg
2311	agttccaata	ccgtatatct	tcttatattg	ttctattgce	tccaatatgt	attcttcgct	taattgtaga
2381	tactcagaca	actcatacaa	gttactgtac	ccataattgt	aagcttctac	aattttcgct	aacgggactg
2451	ctgagataaa	gcggtgtcgt	cttgcgtaat	tttcgaactt	gcgattgttg	aatttcgatt	gacttaaaat
2521	gttgccatac	gtcaacttgt	gggtgggcaag	ttcttcatat	aatacttcta	atttgttctt	ttcggaataag
2591	gaaggtctaa	taaaaaatttc	tccttcttga	taccaacctt	cgaaatcctg	aggtactcct	tgtgtttctt
2661	tcacttcaac	ttcacatttc	ataagcaatt	cttcgtattt	ttccatgcgc	caaaacctt	tggtgtctta
2731	tttctttcta	tctctaaccc	attgcataaa	attttcgatt	tcttccatt	cttcgggagt	aaattcatct
2801	ttattttgat	gaccggttat	agtttcttga	tgaatacttc	tttcttctgt	aattctcgat	ttaggtacat
2871	taaaagtaac	tgctaatgtt	tggtactttg	atattctagg	atatttaagt	tcttttaagc	agttagagat
2941	tgtttagtat	cttaaccgga	ttgcttcaga	caattctact	tgagttaagt	tggttctctt	cataagttgt
3011	tctaaagttct	ctgataaaat	ttttctagca	ctcttatatt	ccataatttt	ctctcttagt	attacttaat
3081	gtaataactaa	tttaccataa	gttaatatcac	ttttcaatac	aaaattattac	ttttttgaaa	taaatatcac
3151	tttaggtgtt	gacataattac	tttaagtgtat	agtatagttg	taaatgtcaa	cgggaggtga	tacgaattgc
3221	cagaaaaattt	taaaagagttc	tctgttaaagg	tctggagaac	taattcgaat	atgacacac	aagatgtcgc
3291	tgataaattta	ggcgttacta	aacaatctgt	aataagatgg	gaaaaagatg	acgcagaatt	aaaaggctta
3361	caattgtatg	ctttagccaa	attattcaac	acagaagttg	attatataaa	ggctaaaaaa	atttaacatt
3431	aatatcactt	taagtataaa	aggaggaaac	tgaatgcaa	gaattacaaa	catttaattt	tgaagaattta
3501	ccagttaagg	aaattgaagt	gggaaggagaa	cccttctttt	taggttaagg	tggtgtctga	atttttaggtt
3571	atgcacgagc	agataaacgc	atacgcacac	atgttgatag	tgaagatagg	ctgatgcacc	aaattagtgc
3641	gtcaggtcaa	aacagaaata	tgatcatcat	caacgaatct	ggattataca	gtttaatctt	tgacgcttct
3711	aaacaaagta	aaaacgaaaa	cattagagaa	accgctagga	aattcaaacg	ctgggttaact	tcggaagttt
3781	taccgacgtt	aagaaaaact	ggtgtcttacc	aagtacctag	tgacccaatg	caagcattga	gattaatgtt
3851	tgaagctaca	gaagaaacaa	aacaagaagt	taaaaacgtg	aaagatgatg	ttattgattt	gaaagaaat
3921	caaaaaactgg	atgcggggaga	ctacaatttc	tttaactagaa	caatcaatca	aagagttagct	catatacaaa
3991	gactacatcg	gatacaaaac	caaaaacacac	gtagcgaaat	attcagggat	attaattcag	aagrgaaaaa
4061	gatgactggt	gcgagttcaa	gaacgaacgt	aagacaaaaa	catttcgacg	atgtaattga	aatgattgct
4131	aatgtgttcc	cgtcacaaagc	tactttatatac	agaatcaagc	aaattgaaat	gaaattttta	aacgaaatat
4201	aggagaggtc	gaatatggaa	tacatcggtat	atgcagacgc	aaatgcgttt	gtaaaaataa	gtggcatttc
4271	aaaagatgat	ctagagaaaa	aagtctactc	gaacaaagag	tttcaaaaag	aatgcattga	cagatttggt
4341	cgaggacaaa	agcgttatat	aaaaattgac	aaagctattc	aatttatcgg	taccaattta	atgattaatg
4411	aatacgaatt	ataggaggag	ttatcaaatg	agtaaaaact	ataaaaagta	cctagttagca	gtactatgct
4481	tcacagctct	agcagttgta	cttatgcctg	ttctatactt	cactacagcg	tggtcaattg	caggattcgc
4551	aaagtatcga	acattcatat	actacaaaga	atacttttat	gaagaataaa	aaaactgcta	cttgcgtcaa
4621	cagtaaacag	tgacaacacat	ttatcaaaat	atacaactta	attaaatcaa	aatatccgga	ggtagtcac
4691	tatggtctgaa	aatattaaaa	ctgaacacaa	ttattacact	aaagatttct	caggatagac	aaatgaagaa

4761 gataactttg tagcaaatca agaattgaca gtaacaatca cattgaacga gtacagaaaa cttattgaaa
4831 taaaagggtgt taaagataaa gaagaagata cttacagagg taagtatttt gcggaagaaa gaaaaaacga
4901 aaaattggaa aaagaaaaata taaaactaaa aaacaaaatt tatgaattac aaaacgaaga agataacgag
4971 gaggacgaag aagacaagga ggacgagaac gatgtattac aaaattggtg agataaaaaa caaaattata
5041 agctttaacg ggtttgaatt taaagtgtct gtgatgaaga gacatgacgg tatcagtata caaatcaagg
5111 atatgaataa tgttccacct aaatcgtttc atgtcataga ttttaagcga ctatatattg cgacggatgc
5181 aatgcgtgac gttataaacg aatggattga aaataacaca gatgaacagg acaaaactaat taacttagtc
5251 atgaaatggt agggaggtatg aaaagtgaat gatttacaag agagagaatt agaaacattc gaacaagacg
5321 accgattcaa actcaactgat ctgacagtg ctaactgggt ttttaagaaa ctggatgcaa tcacaactaa
5391 agagaatgaa atcaacgatt tagcaataaa agaaattgaa cgcataaacg aatggaaaga taagaagta
5461 gaaaaattac agagtggcaa agaattatta caaagccttg taattgaata ttacagaata caaaaagaac
5531 aagatagcaa attcaagtgt aatacacctt acggaaaagt gacagccaga aaaggttcaa aagtcattca
5601 agtttagcaat gagcaagaag tcattaaaca acctgagcaa cgaggttttg acaactatgt aaaagtaact
5671 aaaaaactta gccaatcaga cattaaagaa gatttcaatg taactgaaaa cgggacattg attgacgcaa
5741 acggcggaagt ttagaggggt gctagcattg tggagaacac aacgtctac acggttaaagg tgggagaata
5811 gatgactgaa aaaactaatc aagatgtcga tattttaacg caactagggt taaaagacat cagcaaacaa
5881 aatgcaacaa agttttataa atttgcgata tacggcaagt tcgggtactg taaaactacg tttttaacaa
5951 aagataacaa tcttagta ctgatatata atgaggacgg aacaacggta acagaagatg gggcagttgt
6021 gcagattaa aattataagc atttttagtg agtgattaaa atgctgccta aaattattga acaactaaga
6091 gaaaaacgaa aacaaattga tgtttagtg attgaaacaa tccaaaagtt acgtgatatc actatggacg
6161 acatcatgga cggtaaatca aagaacccga catttaatga ttggggcgag tgtgctacac gcattgttag
6231 tattttatcgt tatattttcta aattacaaga acattatcaa tttcatcttg ctataagcgg acacgagggc
6301 attaacaaga acaaagatga tggagggagt actatcaatc caacatcac gatagaggca caagaccaaa
6371 taaaaaaagc agcatcagt caatctgacg tgttagcaag aatgacaata gaagaacatg agcaagcgg
6441 cgaaaaaact tatcaatatg tacttaacgc tgaacatca aatttattcg agacaaagat aagacactca
6511 agcaacatca aaatttaacaa caaacgtttc attaatccaa gtattaacga tgttgtacaa gcaattagaa
6581 atggttaata aaaaataatt aaagggacgg tataaaaatt atgaaaaatca ctggtagaac acaatacatt
6651 caagaaacta atcaagaggg attcatgaaa ggtggggact ttttaggagc tggagaattt acagtataag
6721 tgcgaatgt cgagtttaac gacagagaaa acagatactt cagcatgtgt tttgaaaaa acgaaggtaa
6791 acaatacaaa cacaacaaat tgcgtccacc attccaacaa gattatcaag aaaaacaata tatcgagta
6861 cttagtagat taggaattaa attgaactta ccagatttaa cttttgacac agatcaatta attaacaana
6931 tcggaactat tgtacttaaa aataaattta acgaggaaca aggcagat tttgttaagac tctcatatgt
7001 aaaaatttgg aataaagacg atgaagtatg taataaacca gaacctaaaa ctgatgagat gaaacaaaaa
7071 gaacagcaag caaatggtta acagacacct atgagtcaac aatcaaaccc attcgtaat gctaattggtc
7141 caatagaaat caatgatgat gatttaccgt tctaggacgt ggtttaaatg caatacatta caagatacca
7211 gaaagacaat agcgttactt attcgcgtgt tgcactggtt gttgaacttg aacaaagtca cattgtatta
7281 ctgaaaaacg gatattccgct aaaagcagaa gttagaggtt cggacaataa aaaactatct atagaacaac
7351 cgaaaaaatt attcgcaatg tgtagagata tagaacttca ctggggcgaa ccagtagaat caactagaaa
7421 attattacaa acagaaattg aaattatgaa aggttatgaa gaaatcagtc tgcgtgactg ttaagttaa
7491 gttgcgagag agttaataga actgattata tctgttatgt ttcacatca aatacctatg agtgtagaaa
7561 cgagttaagt ttaagcgaa gataaagcgt tatttatatt ggttacaatc aaccgcaact gtgtaatatg
7631 cggaaagcct cagcgagacc tggcagatca tgaagcagtc ggcagaggtta tgaacagaaa caagatgaat
7701 cactacgaca aacatgtgtt agcactgtgt agacaacatc ataataaca gcacgcaatt ggtgttaagt
7771 cggttgatga taatatcaa ttgcatgact cgtggataaa agttgatgag aggttcaata aaatggttaa
7841 aggagagaaa aatgaataag ttactaatag atgactatcc gatcacagta ttaccgaaat tagctgaaat
7911 aataggggtta aacgaagcaa tagtattgca acaaatccat tattggctaa acaactcaaa acataaatc
7981 gatggcaaaa cttggatttt taattcttat ccgaatggc aaaaacaatt tccatttttg agcgagagaa
8051 ctataaaaaa gatatttggg agtttagaaa attgcatgta ggttaactca acaaggtctgg
8121 atttgaccgt acaaaatggt attcaatcaa ttatgaacaa ttaaacaaac tagtggcagc accatcgagg
8191 aaaaatggcc cgacgatgag gacaaattgg cagcatgcaa gaggacaaaa tgaccggacc aataccattg
8261 actacacaga actaacaaa catagagaga cagacgacgt ctcaaaagtc ttaagtata ttagtaccaa
8331 tttagaattt atacaaaacc ctttaaaagc agaacagtta gaacacgaaa ttaaatcatt taagcaagat
8401 cagtctgaaa tagtaaaagt agctaccgat tactgcaag aaaaacaaca aggtctgaat tacttactaa
8471 ctgtattaaa gaactggaaat aatagaaagg cttcagataa agaaaagtgt gaaaacaaat tgaacactcg
8541 taactctaaa aaagaaacta ctgatgatgt catagcacia atggaaaaag aattgagtga tgactaatgc
8611 cgatgagcaa aacacaagca ttgaaattta ttaaaaaagt taggtacgta tacaacatcg attttgataa
8681 accaaagtta gaaatgtgga ttgatgtatt aagtcaaaac ggggattatc aaccaactgt aaaagctgta
8751 gatggatata tcaacagtaa caaccgctac cgcctaaccc taccagcaat catgctgaag gcacctaaaa
8821 aagttatctat tgagccggtta gacaacgaaa ccgctacaca ccaatggaaa atgcagaatg accccgaata
8891 tgtcagacaa agaaaaaatag cgctagataa cttcatgaat aagttggcag aatttggggg cgataacgaa
8961 tgaattacgg tcaatttgaa attgaaagca caataatcgc tacgctactt aaaaacacgg acgtactaga
9031 aaagataaga gttaaagatt acatgtttac gaacgaaaag ttttaaacct ttttcaatta tgtaattgac
9101 ttcggaaaga tagatcatca agaaatctat ttaaaagcaa ctaaaagataa agagttttta gatgcagata
9171 ctataactaa actttacaac tccgatttca ttggatagcg attctttgaa cgttatcaac aagaattatt
9241 gtaaatgtta caaatcaaca aagcgaagaa attggtaact gagttcaaac aacaacctac gaacaaaaat
9311 ttttaatact ttagtgatga actcaaggat ttataaaacaa ttactaacag aaaaagaagc ggaacccaaga
9381 agtttgttga gtagttgtgc gatgagttat acagcgatag ccctaagaag caaattaaga cgggttatag
9451 gctctgggat tacaataatg ggggattgga gccgtcgcaa ttaactgta ctcgcagcgg tccctcagtg
9521 ggttaagacg gttttgcatt aaacatgatg ctgaacatag cacaataatg atacaaaaca tctttcttta
9591 gtctcgaaac aactggcaca tcagtattga aacgtatggt atcaacaatt actggtattg agttaacaaa
9661 gatataagaa atcaggaact taacggcggg tgacttaaca aagtttaacga atgcgatgga taaaatcatg
9731 aaattagcaa tctgatattc tgataaaaag aatatcacac cgcaagatgt cgcagcgcaa gcaattgagg
9801 attcagacag gcaacaaggt atttttatag attatcttca actgatggat actgatgcga aagttgatag
9871 acgtgtagca gtagaaaaga tatcacgtga cttaaagata atcgctaacg agacagggcg aatcatcgta
9941 ctactttcac aactgaatcg tgggtgtcag tctagacagg ataaaagacc aatgctatcg gacatgaaag
10011 aatcagggcg aatagaagca gatgcgagtt tagcgatgct actttaccgt gatgattatt ataaccgga

10081 cgaagatgac agtatcactg gcaaatctat tgttgaaatgt aacatagcca aaaacaaaga cggcgaaacc
 10151 ggaataattg aatttgagta ttacaagaag actcagaggt ttttcacatg aatataatgc aattcaaaag
 10221 cttattgaaa tcgatgtatg aagagacaaa gcaaacgcgac ccgattgtag caaatgtata tatcgagact
 10291 ggttggcgcg tcaatagatt gttggacaat aacgagttat cgcctttcga tgaattacgac agagttgaaa
 10361 agaaaatcat gaattgaatc aactgggaaga aaacacacat taaggagtggt taaaaaatgc cgaagaaaa
 10431 atattactta taccgagaag atggcacgga agatattaaag gtcatcaagt ataaagacaa cgtaaatgaa
 10501 gtttattcgc tcacaggagc ccattttcagc gacgaaaaa aaattatgac tgaatagtag ctaaaacgat
 10571 ttaaaaggcgc tcacgggctt ctatatgagc aagagctagg attgcaagca acgatatttg atatttagag
 10641 gtggcacaat gagtaaatat aatgttaaga aagttgagta caaagggaatt gtatttgata gcaaaagtaga
 10711 gtgcgaatat taccatatt tagaaagtaa tatgaatggc actaactatg atcgatcgga aatacaaccg
 10781 aaatttgaat tacaacctaa attcgggaaa caaagaccga ttacgtatat agccgatttc tctttgtgga
 10851 aggaaggga aactggttga gttatagacg ttaaaggtaa ggogactgaa gttgccaaca tcaaaagcgaa
 10921 gatattcaga tatcagtaga gagatgtgaa ttttaacgtg atatgtaaa ggcctaaata cacaggtcaa
 10991 gaaatggatg tatatgagga cttagtga aaagaaaag gtcagacgta aagaaatgaa tgaatctaat
 11061 caacaacaa catatataaa cgcacaatt gatataaaga taactacaga agttgaatat cagcattacg
 11131 atgatgtgga taaagaaaaa gatacgtg caaagcgctt agatgacaa cgggacgaat tactaaagta
 11201 tgacaacata acaataaagc atgcataat agaggttgaa taaatgaagt tgaacgaagt attcgcaact
 11271 aatttaaggg taactcatgg tagagataac gtaagtgtcc aagatttgca caatgaaact ggcgtatcaa
 11341 gatcaactat tagtggatat aaaaacggaa aagctgagat gggttaacta aatgtattag ataaattggc
 11411 agatgctcta ggtgttaag taagtgaact atttactaga aatcacaa caacgcaaaat agaggattgg
 11481 attaaaaaag taaatgtata gaggtggaa ataatgagat cgttaagatt aacggtaaac catataaatt
 11551 taccgaacat gaaaatgaat tgataaaaaa gaacgggtta actccaggaa tggttgcaaa aagagtagca
 11621 ggtggctggg cgtgtttaga agccttcat gacccctatg gtatgcgctt agctgagtat aaagaaattg
 11691 tegtattcaa aatcatggag cgagagagca aagagcgtag aatggttagg caacgacgta aagaggctga
 11761 actacgttaa aagaagccac atttgtttaa tgtgctcaa aaacattctc gtgactcgtg cttgttcgat
 11831 gtcactcata acccaactgt caagaaatgg agtgaagcat aatgagcata atcagtaaca gaaagtaga
 11901 tatgaacaaa acgcaagaca atgttaaa caacggcgcat tacacatagc gcaacttga aattatagat
 11971 tttatcgaa aggttacggc acagtatcca cctcaactag cattcgcaat aggttaatgca atcaataact
 12041 tgtcttagag accgttaaa aatggatcat aggatttag aaaggcgaa ttttactgct aaagagcttt
 12111 tgacttgtgg gagggttaac gatggcaacg caaaaacaa ttgattacgt aatgtcatta caggaaacat
 12181 tgggattaga agactgtgaa aatatatagc acgaacaagt taaagctatg agtcataaag aagtagcaa
 12251 tgtgattgaa aactataaga caagcatatg ggtgaagag ctatataacg aatgcatgtc gttttgctg
 12321 cctaattgct aaaaaggagt atgacccatga acgatagcgc acgcaaaaga tacttaaac aattttcag
 12391 ctctaagaga tatctgtatc aagacaacga gcgagtgcca catatccatg tagtaaatgg cacttattac
 12461 tttccaggac attataaaac gatgttttaa ggcgtgaaaa agacatttga tactgtgtaa gagctcgaa
 12531 tatatacaaa gcaacatgat ttggaatatg aggaacagaa gcaaccaact ttattttaga ggaagtggaa
 12601 ataattggcaa agattaaaag aaaaagaag atgacgctac tcgaactggg ggaatgggca tggacaacac
 12671 ctgaacaagt tgaagtaaa gtgtttcaat cagatagaat gggcacgctt gggaatgta cogaagtaca
 12741 tttttcaact gatgggcatg ggttttatc aasagtagta acagataaag atatttttac tgaagaatc
 12811 acagaggaag tcaactgaaga tactgagttt gattgtctag tagaaactaa cgaattgaa ggttttgaa
 12881 tatatgaaaa tgattcaatc agagagttga tagacggtac ttccagagcg ttttatatac taaacgaaga
 12951 taaaactarg acartaattt ggaagatgg ggaagttgta gtatgatgca aacctataaa gtatgtctt
 13021 gtatcaattt ctttgcattc aatgtgatt ataaatttaa gaaacattat ttctgtgaaa gtcagaatga
 13091 agaaaagccc acgaacatgg tatttaaaat gattcgtaaa aagctcccggt tcgaactgca aagcatgaa
 13161 gtcgaaaaag tggaggcaat ataatgatc aaccaacaa aggaagatta attaatcca tgaaaaaa
 13231 tggagctgaa aatgttgact ctatcactga tgagcaaat gcaataaagc actttagagc tcaatcaaaa
 13301 gttttttaa agcaacgtga tgagataaag aagcaacgag atgagcttat cgaaggatata gctaagttaa
 13371 gaaaacgtaa cgaagagctg gagaacatgt ggcgcacagt caaaaatgaa ttgcttgtaa gatacgaaca
 13441 ttactgtttt aaaaattag aactacaccc tgagagcaaa gcgaacagga taggagctct ctatatgga
 13511 ggtaaaaaga ctgcagatat tatactgtcg cgaatggaag aactagacgg aacaaatgag ttctacgaat
 13581 ttttagggca aatggaggca gacacaaatg aataacccgt aacaattaga acaatcagtg atcagtaata
 13651 gtgcgtataa cgttaatgac acagaggggt tactaaaaa gattgaggac gtgtataaga aagcgaagc
 13721 gtttgatgaa atacttgagg gaatgacaaa tgctattcaa cattcagtta aagaaggtat tgaacttgat
 13791 gaagcagtag ggtttatggc aggtcaagtt gtcataaat atgaggagga gcaggaaaaat gaggattagt
 13861 taggagata aagtatataa ccatgaaaca aacgaaagtc tagagattgt gcaattggct ggagattat
 13931 gagatacaca ttataaactg tctgatgatt cagttattag cattatagat ttattacta aaccaattta
 14001 tctaattaa ggggacgagt gagtggatg gaaacgatta aaaaatgtgg tgcgcacccc agttatcaaa
 14071 aataaaaatt taaagtcggt atacgttaaca aaagataatg tgaagaggt tcaaaaagaa ttaggtttct
 14141 ttgaaatttt taatgaagaa gtgttattaa ctggattttt atcatttcaa aggataccta ttacattat
 14211 ttggattaat cctaactctc ataaagacgc tagatattac ttgctaaagc agctagagat tgaagatat
 14281 tttgaatttt tggaggagca gtaaatgctt gaaatcatcg accaacgtga tgcattgcta gaagaaaagt
 14351 atttaaacga cgaactgtgg tacgagctag attattggtt gaataaacgc aagtcagaaa atgaacagat
 14421 tgatattgat agagtgttga aatttattga ggaattaaaa cgaataggaga taacgaataa atgaataatt
 14491 taacagtaga tcaattaaaa gaacttttac aaatacaaaa ggaagttcgac gatagaacac cgaatagaaa
 14561 ttttaattgac acagtagcta gtatgattat tgaatttcg gagtgggtta acacacttga gttttttaa
 14631 aattggaa gaacaaccagg taagccatta gatacaaat tagatgagat tgctgattac tttagtttca
 14701 gtttgcaatt aactctgact attgttgatg aagaagattt ggaagagact actgaggtta tgggtgattt
 14771 gattgaaat gaagtactt taactaaact acattcagtt tattttgttc atgtaatgca tacactaaca
 14841 gaacaatttg taaaagggtt tgataaatgt attgtacaag ttttaataat gccttttttg tacgccaata
 14911 cttactatca aatcgaccaa ctcatgacg catacaaaaa gaaaatgaaa aggaaccacg aaagacaaga
 14981 tggaaacgca gacgcaggaa aagatcacgt ctaagacat cttagatcga gtaacaggag ttttggggaa
 15051 gtgacgaat acttagtcac aacattcaaa gattcaacag gacaaccaca tgaacatttt actgctgcta
 15121 gagataatca gacgtttaca gttgttgagg cggagagtaa agaaggagcy aaagagaagt acgagaacaa
 15191 agtttaagata agggagatg gagatggcaa agaaaacggt aacgatgtat gtagatgaaa acttattagt
 15261 agtagctagt aatgaaatat cagaactatt atatgaatat gacagtgagt taatgtcagc tgaatgaagt
 15331 ggcgataata gagatatcga aaaaaaaga gacgattaa aacagcctac acaaatatc gataaattaa

15401 catgtctgagg aggcagacga tgattaacat acctaaaaatg aaattcccga aaaagtacac tgaataaatc
15471 aagaataata aaaataaaac acctgaagaa aaagctaaaga ttgaagatga ttctattaaa gaaactaatg
15541 ataagacag tgaattttac agtccctatga tggcctaatac gaataaacat gaactaagggt ctatgttaag
15611 aatgatgcct agtttaattg atactggaga tggcaatgat gattaaaaaa cttaaaaata tggattgggt
15681 cगतatctttt attgtctgaa tactcgcat attcggcgta atcgccactga tgcctgtgtgt catatcgctt
15751 atctatacag tggctagcta ccaaaacaaa gaagtatac aagggacaat tacagataaa tataacaaaga
15821 gacaagataa agaagacaag ttctatatgt tcttagacaa caagcaagtc atcgaaaact ctgacttact
15891 attcaaaaag aaatttgata gcgcagacat acaagctragg ttaaaagtga gcgacaaagt agaagttaaa
15961 acgattgggt atagaataca ctttttaaat ttatatccgg tcttatacga agtaaaagaag gtatataaat
16031 aatgattaaa caaatattaa gactattatt cttactagcg atgtatgagc taggttaagta tctaaactgag
16101 aaagtatata ttatgacgac ggctaattgat gatgtagagg cgcggagtga cttcgcaaaag ttgagcgatc
16171 agtctgattt gatgaggcg gagggtgcag agtagatgta tagcaaaagag tcaattgtta atatgataag
16241 cacacataaa atgaagtga atgtattagc tgatgtaata ccggaatatg atagcaattc aattgcacag
16311 tatggcatat aagcaacggt gccgaaacca aaggggaaa actcaagtga agttgaagat gttgttgtga
16381 ggcttgagag agcaaataaa aggtatgctc agatgttaaa agaggttgag tttataaatc aatcgcaaca
16451 gagattggga caggttgact ttgtctctt agagtattg aagaaaaggt ataacagggga tgcgattatc
15521 aagaagatgc ctaactctaa attaaataga aacaactctt tagcgccggc tgatgagttt gcgaaaaaga
15591 tttatctact acagtacga aatgacaaa aatgacagaa atgacgaaa tgacactatt ttaactctgt
15661 gaatttaatt tatataattg atttgaaga attatcttaa gacgtgggg aatagccaca ttagatgttc
15731 tcatcgatgt gattgagaag tgacaacat ataaaagatg atatgttacg ctattaatca cctactacct
15801 gccatataat tgggtagttt aattcttgca ttttgagta taactatttt cctcctttca cattttattga
15871 acgtagctcc tgcacaaagt gtaggggcat tttttatatt taataacta gagtaattaa cgtaaagggc
15941 tgtgatacag tgaaaaacat tgattaaatt aacaccgaag caagaaaagt ttgtgttagg acctatagag
17011 ggcaagagcc aacggaaagc atatatgac gcaggggtatt cgactaaaag taagagtggg gaatatctag
17081 ataagaaagc gagtacact tttaaaaatc ggaaggtttc cggaaaggtac gaaaaattgc gtcaagaagt
17151 agctgaacaa tcaaaatgga cagcccaaaa ggctcttgaa gaattatgag ggctaaagaa tgtacttaag
17221 aatgacattg aaatagagg agtgaagaaa gcgacagctg atgcattctc cgctagttta gatggatga
17291 atagaaatgc gtttaggtaac gaagttttay ctaaaaagaa aatagaacct gaaattaaaga tgccttgaga
17361 gaagattgaa caaatagata aaggtgacag tggaaacagaa gataaaatca aacaacttca cgcagcaata
17431 acggaagtga tctgcaatga ataaacttaa atctttatat acggaacaaa aatttgaat attgaagcaa
17501 acgcaaaaaa aagattgggt tatgttaatt aatcacggag caaagcgta aggtaaaaa atattaaaca
17571 atgacttatt tttaactgag ttatcgctg tgcgaagat agcagacgaa gaaggaaattg agacacctca
17641 atataatactt gctggctcaa cattaggtac gattcaaaaa aacgtactaa tagagttaac taacaaatat
17711 ggcatgtagt ttaattttga taaatataat tcatctatgt tatttggcgt tcaagtgggt cagacaggtc
17781 acagttaagt aagtggtata ggagctatag gtgggtatgac atcgtttgtt gcatatatca atgaagcgtc
17851 gttagcgcat gaagaggtgt ttgacgagat taagtcaact tgttagtgaa ctggtgcaag aatatgtgta
17921 ctacccaacc ctagccatcc cagacattgg ttgttgaaag attatattga aaatacagat cctaaagcag
17991 gtatactgag tcaaccaattt aagctcgatg acaataactt tcttaattgat agatataaag agtctattaa
18061 ggcttcaaaa ccatcaggtt tgttctatga acgttaatat aacgggtatgt ggggtgtctg tgacggtgta
18131 gctatgcggc accttgattt gaactgaat acgattaaag cagatgaact ggacgacata cctatcaag
18201 aatcctttgc tgggtgcag tgggggttac agcactatgg atctatttgt ttaattaggac gaggtataga
18271 tggtaacttt tattttattg aggaagcagc acaccaattt aagtttatgt atgattgggt ggttattgca
18341 aaagataatt taagtataga tggcaatatt aatttttact gcgatactgc acgacctgaa tacatcactg
18411 aatttagaag acatagatta cgtgcaattt acgctgaraa aagtaaaacta tccgggtgtg aggaagtgc
18481 caagtgttc aaacaaaaca agttacttgt tctttatgat aatatggata ggttttaaga agaggtattt
18551 aaatatgttt ggcacccctac aaacggagag cctataaaa aattttatga cgtgttgagc tcyttaagat
18621 atgcacatac cacacatact aaacctgaac gattaaaggg ggggaaatga cattgtataa gtttaatatg
18691 gatattgaag caataaggat attgcctaag catattgagg ctctaataga gtcacataaa gacgataag
18761 agagaatgggt taactcttat aatagatata agacacatat tgactatgta ccaatatca aacgtcgacc
18831 aattgaagaa aaagaagatt ttgaactgg tggaaatgta aggcgattag acgtgtcgt taataacaaa
18901 cttacaact cttttgacag cgaatttgtt gatcacagtg ttgggttatt acatgggtgt cctgttactt
18971 atgatttaga tgaaaacgca gaaaaaacg aaaaagttga aagttttata accaactttg cctattagaaa
19041 tagtgttgat gatgaggatt ctgaattagg taaaatggca gcaatttgct gatattgtgc taggttagca
19121 tatattgata cgaatgggtg tattaggatt aagaatatag atccctataa tgttattttt gttggcgaca
19181 atattttaga acctacatac tcaattgcgt acttttatga aaaagatgat gataatggca ctgattatgt
19251 gtacgcagag ttttacgata atgttatta ttatgtattt cgaggagaag gtattgacgc tttgcaagaa
19321 gttggagcat atgaacattt atttgattac aatccattgt ttggtgtacc taacaaacaa gagatgatag
19391 gagatgctga aaaggttatt cacttaattg acgcatatga ttaacaatg agcgtatgat caagtgaat
19461 tagtcagaca cgtttagcat accttgtgtt acgcggtatg ggtatgagtg aagaatgat tcaagaacaa
19531 caaaagagtg gcgcatttga gttgttcgac aaagatatgy acgttaataa cttaaacaaa gatgtaaatg
19601 acacaatgat tgagaacctt ttatagcgaa tggaaaagaa tatcatgcyt tttgcaagat cagtaaatc
19671 taattctgac gagtttaacc gaaatgtacc tatcatttga atgaactta aacttatggc tttagagaac
19741 aagtgatga cgtttgagcg taagatgaca gctatgttga ggtatcaatt caaagttatt ttatctgcat
19811 taaagcgtaa aggttacaac ttggatgatg atagttattt aaactgtata ttaagttca ctcgtaacat
19881 tccagtttaa aagttagaag aatcacaagt gctaattaac ctgaaggagc aagtttcaga acgaacagg
19951 ttaggacaat cacaactagt tgatgatgtt gattacgaat tagacgaat ggaaaaagaa agtcttgaat
20021 ttaatgacaa attacctgac atagatgaag gtgacgcaaa tgacaatacc caaataaacc aatcagaatg
20091 atattgata gatatcgag ggttataatc ctaaaagcaga aaaaacataa gaacaactat ttgctaactg
20161 acctaaagag ataaaaacaa tcatcgaga tatgttttag aatatcaaa atgatgatgt gtatgttaca
20231 tggactgaat tcaataaata caacaggctc aataaggagt taactcgtat aggtacaatg ttgacttatg
20301 acctaggca agtagctaa agtattcaga agtcacaaga agatgcttat atagaaaaat tcttatgag
20371 cctttattta tatgaattgg caggtcaaac atctatgcag ttgtatgttc cgagttaaga ggttaataa
20441 ttagctattt acaaacctat tgagttcatt cgttttaatg caacactaca aaaaacatcg gatgaagtat
20511 tgaaaaagat acgtatgcac attacacaa gtattatgag tggagagggg tactctaaga tagctaaagc
20581 aatagctgat gatgtcgga tgcctaaagc tcaatcattg cgtgtggctc gtacagaagc aggcagagca
20651 atgtcacaag cttggacttga tagcgcaatg gttgctaaag ataacgggtt gaatatgaag aaactgtggc

20721	atgctactaa	agatacacga	acacgtgata	ctcatcgtca	tttagatggg	gaatcagtg	aaatagatca
20791	gaatttttaa	tcaagtgggt	gtgttgggca	ggcgcccaag	ctatttattg	gtgtaaacag	tgcgaagag
20861	aatattaatt	gtcgttgcaa	attactttat	tattattgat	aaaatgaatt	gccaaactga	atgagagcac
20931	gtaaagacga	tggtaaaaat	gaagttatcc	cattcatgac	ttatcgtgag	tgggagaaat	ataagcgaaa
21001	aggtcggtat	tgataatgat	tttaaaataa	aagtaaatgt	tgataactgg	gaagctatag	aaaagttaga
21071	acgcattaaa	tccttgcacg	aagagataat	agagttacaa	aacgaaaaag	ttgttgtaaa	cgtaacagtt
21141	aaaaatgaag	ctgattttag	tatggttaaa	acatctatta	gcgaagaaaa	tgcataaaat	aatgatttca
21211	cactttttta	gttgtctctt	tgctactcga	ccttagcatg	tcgttaaaat	gctttttatt	atgcactttt
21281	cggactgtta	gggtacgcga	agggcaaaaa	ggagttttga	tatatgaata	tcgaagaagt	taagtctttt
21351	tttgaagaac	acaaagacga	taaaagaagta	aaagattatc	taaagggact	taagacgggt	tctgttgatg
21421	acgttaaaag	cttttttagat	acagaagaag	gtaaacgatt	cattcaacct	gaattagatc	gttatcattc
21491	gaaaggatta	gaatcatgga	aagagaaaaa	tcttgaggat	ctaactgaac	aagaagtacg	gaagcgtaat
21561	cctgagcaat	cagaagaaca	aaaacgtatt	agtgtctctg	aacaagaggt	agaaaaacgc	gacgcagagg
21631	caaaacgtga	gaagtttaaga	agtaaacgcg	taggttaaac	gcaggaacta	aatctaccaa	catccttagt
21701	tgatagattt	ttaggcgatt	ctgatgaaga	tactgagcaa	aaacttaaa	ctttaaaaga	aacctttgac
21771	aaagtattgt	aaaaagggct	tgagtctaaa	tttaaatcga	gtggaagaga	tgtaaaagaa	tcacgaatc
21841	aaagtattga	cccttcaaat	gtaaagtcca	ttgaagaaat	ggcgaaagaa	atcaatatta	gaaaaataag
21911	ttaggttaata	aaatatggca	actccaacat	acacgccagg	caatgttatt	ttatcggatt	ttaaaaacgg
21981	cggtattcca	gcagaacaag	gtactttaat	catgaaagac	attatggcta	attcagcaat	tatgaaatta
22051	gctaaaaaat	agccaatgac	agcacaaaag	aaaaaattta	cttacttagc	aaaaggtgta	ggcgccctact
22121	gggtatcaga	aacggaacgt	attcaaaact	ctaagcctga	atatgcgcaa	gcagaataag	aaagctaagaa
22191	aatgggtgta	attattccgt	tatcaaaaaga	gtttcttaaa	tggactgcga	aaagtattct	taatgaggtt
22261	aaacctctaa	tgcgagagcg	attttacaaa	gcgtttgacc	aaagtgttat	ctttgggtact	aaatcacctt
22331	acaacacttc	aaactagtgt	aaaacgcgtg	ttgaagggcg	agaagagaaa	ggtaacgttg	ttacagatc
22401	taataattta	tacgtagacc	tttcggcatt	aatggctact	attgaagatg	aaagtattga	ttcaaacgga
22471	gtatttaact	caggttcatt	cagaagtaaa	atgcgtaatg	cttttagatg	taatgacaga	ccattatttg
22541	atgctaacgt	gaacgagatt	atgggtattac	caactatctt	tactggagcg	gatgtatagc	acaaaaagaa
22611	atcgttagca	ctaactgggt	attgggtatta	cgcaacgttac	ggatctctac	aaagtattga	gtatgcaatt
22681	tctgaagatg	ccacgttaac	gacgtttaca	gcacagatg	cttctggcca	accagtatca	ttattgaaac
22751	gtgatattgt	cgctttacgt	gcgacgatgc	atattgcata	catgaacgtt	aaaccagaa	cggtcgcaac
22821	gcttaaacca	actgaatagg	aggagatatg	atggctaatc	ctgcagagaa	gattaaggtta	aaaaaagaca
22891	atatgactat	tactgtttaca	aagaaggcat	ttgactctta	ttacagtctt	gtcggttaca	aagaggttaa
22961	atcacgtcgt	actacgtctg	ataagagcga	gtgataaaaa	tgactcttta	tgaagatgtt	aaacttttac
23031	tcaagaaaaa	tggagtggaa	gttaaaagtg	atgaagaaga	aatatttaag	atggaaagtg	acggaatact
23101	agaagatgtt	agggatatata	caaacataga	ttttatgaaa	gatgggtcaag	tcatttatcc	ttactcaatc
23171	aaaaagtagt	tcgcatagtg	cctagagtat	tatcaacgac	ctgaagttaa	aaagaattta	aaagtcaagaa
23241	gtatcgggac	agtgctcgac	acttataacg	atgggtgtccc	tgattacatt	agtggagtat	taaacaggtta
23311	taaacgagca	aaagtctatc	cggttaaac	aaataaggtag	aggtgtgtgt	tggttttaac	ccatcacgag
23381	aatccctcca	cactattttt	attgggaagta	tcaaaaaagt	aggagagtat	ccaattatcc	aaagcgcttt
23451	tgttaagcgt	aaaacaaatta	aaagatttat	ggatacgcct	actacatctg	aaacactaaa	atttccatcaa
23521	atgtcaacag	aatatgacag	aaacctatat	gtaccttatg	acttgccaat	atctaaaaac	aaatttatctg
23591	agtatgaggg	tagaatcttt	agtattgaag	gtgattctgt	agatcagggc	ggacaacatg	aaattaaagt
23661	actacgactt	aaagcaggtg	catatggcaa	aaagttaagta	cggtgtctgt	agcatgggtg	ttgaattgga
23731	ttaagtctgt	aagaaaaatag	aaagtggtgt	taaaaaaagt	attgctaaaa	caacgacgaa	gatttcaaac
23801	actgctgtag	cattagctcc	tggtgactta	ggttttttag	aagaaagtat	tgacttttaa	tatttcgatg
23871	tggtgttatc	cagtggtata	agtgctcggt	cagattatgc	aatatacgtt	gaatacggta	ctggtatata
23941	tgtctactgt	cctggtggta	gtcgtgtcac	aaagattccg	tggagtttta	aaaggtatga	cggcgaaatg
24011	tacaccacat	atgggtcaag	gccacagcca	ttttggaaac	ctgcaattga	cgacgagcgc	aaagacattcg
24081	agcagtat	ttcatagagg	tggttaataa	tggtgggtatc	agttgagcgt	gaacttaca	atcaaatata
24151	taaaagatta	atctcagacc	ctaacactaa	caaaactagt	gatgataggg	tttttgacgt	tggtcaagat
24221	gacgtgtgtt	acccatata	tggtgtgggt	gaatcaaacg	tcacttaacaa	cgaatctagc	gcaacaaatga
24291	gagaacagct	cggatttgtc	atacatgtgt	atccacagtt	cgctacacaa	tacgagggta	agctcatttt
24361	aagcgcgata	gggttatgtg	ttaacagacc	tatagaaata	gataattacg	agtttcaagt	tagccgtatc
24431	gatagtcagg	cagttatccc	tgatatagac	aggtttacta	agcaatggac	gatacgggtt	ttatttaagt
24501	acagacataa	aaagaaaaac	gaaggagtgt	attaaatggc	gcaaaaaaac	tatttagcag	ttgtacgtcc
24571	agctgaacct	gacttagatc	cagtagaatc	tttattatta	gctgacttac	aaagaaggtg	acatacagatt
24641	gaaatgattt	tagctgaat	agtagcaggg	ggtaaaacgg	actattctcc	caatgcaatg	tcagaatcat
24711	ttaaattaac	aattggtaat	gtgcctggag	ataaagggaat	tgaagcagtg	aaacacgctg	tacaaacagg
24781	tggacagttg	cgtatatggc	tttatgagcg	taataaacgt	gcagacggta	aaacatcagc	aatgtttggg
24851	tatgtgtgtc	cagaatcatt	tgaatgttca	tttgatgatg	aaagtgcaca	aatcgaacta	tcattaaaaa
24921	ttaaatggaa	tacagcagaa	ggtgctgaag	ataacttgcc	gaaagagtg	tttgaagctg	caggtgcgcc
24991	tacagttgaa	tacgaaaaat	tcggcgaaaa	agtcggaaca	ttcgagaatc	aaaagaaagc	tagtgtgtga
25061	cttgattcac	acacggaaga	ccattcttat	taaactaata	gatcaagggg	gcgttaagct	cctatttttt
25131	tataaaaaaa	ttgaaaagag	gtatatattt	tgactgaatt	taatccaatt	acaacattaa	aaattaaatga
25201	cggagaaaaa	gattacgaag	tagaagcaaa	agtaacattt	gcatttgacc	gaaaagctga	aaaattctca
25271	gaagatagcg	aaagtgggag	aaaaggagca	atgccaggat	tcaattgttat	cttttaacggt	ttgctagaat
25341	ctagaaacaa	agcgattttta	caattttggg	aatgtgtctac	tgcttattta	aaaaaccac	caactcgaga
25411	acaattagaa	aaagcaattg	atgattttcat	cactgaaaac	gaggataactt	tgccgttatt	acaaggggct
25481	ttggacaaac	ttacaacatag	tggttttttc	aaaggggaga	gtcgtctcga	ctggatgaca	ttgaacaaag
25551	caccgaatat	ggccaaaagc	gaggacaaag	aaatgacgaa	agcagggata	gaatgatga	aaagaaatta
25621	caaggaatcc	atgggcgcag	aaccttacac	gattactcaa	aaataaggca	actgacagct	agatatttag
25691	gatataatcc	tgaacatgaa	ttgttagcac	taaacacctgc	tgaatggcgt	gattggctta	ttggtggtca
25761	ggataggtac	ctagatcaaa	gacaaattat	aaattgaacaa	gcgcgaagta	acggcttagt	acaagcttct
25831	aagaggtctaa	ctagtatgat	tcgtgacatt	gagaaacaac	gttacgaaat	aagagaacct	ggtagctatg
25901	ctcgtgtaca	aaaagctaga	ttagaagaag	aaaaaagaag	acgtgaaact	ttcaaaagaa	gtacaagaaa
25971	attccttgaa	tcgaagaggag	gttagccttt	ggatactcat	tttatggcaa	agattatggc	caatattaga

26041 gatttccaaa gcaacgtaag gaaagctcaa cgattagcaa agacgtctgt accaaacgaa attgaaacag
26111 atgtaaaaagc agatatttca agattccaaa gagctttaca acgcgctcaa tcaacggctc aacgatggcg
26181 agagcattctt gttaaattat tcatgaaaac agatgagat aaagcgaaat tagaacgcgc taaagctcaa
26251 gttagagcgat ttaacaaca taaagtagat ttgaaactaa gtaacactga attaatggcc aaatataatg
26321 caactaaagc tactgtcgaa gcttgagaa aacatgttgt taagttggat ttagatgcaa accccgctaa
26391 aatggcggtt aagggttcta aagaagattt aatagatctt agcaggtata gttttgatat tgattccagc
26461 agatggaaat taggaaataa attcacaanaa gaattcaatg aagtcgaagg agcaggtaaa cgttcttctg
26531 gaagaattgg tcagattatg agaaaagaag taaatggaa aagtgatatt tggggtaaac ttaacaactc
26601 attgaaagat tacggcgaga aatgggacgc cttagctact aaaatccgaa ctttcggtac tatcttcgcy
26671 caacaggtca aaggcttaat gattgctagt atacaagcat tgataccagt gattgocgga ttagtacctg
26741 caataatggc agtacttaat gcggttggtg tattaggtgg tggcggttta ggtttagtgg gcgcattctc
26811 tgtcgaggtt cttggagtgt ttggctttgg tgcaatggct attagcgctc ttaaaatggt tgaagatgga
26881 acattggcag taacaaaaga agttcaaaac tttagagatg cgagcgatca gttaaaaact acatggcggt
26951 atattgttaa agagaatcaa gcaagtatct ttaatgcgat gtcagcaggt atcagaggcg ttacaagtgc
27021 gatgtctcaa taaaaccat tcttatccga agtatctatg ctagttagag caaacgcacg cgaagttagg
27091 aattgggtta aacattccga aacagctaa aagcggttg aagcattgaa tagcataggt ggcgcaatct
27161 tcggagattt attgaacgct gcaggacgat ttggcgacgg attagttaac attttctact aattaatgce
27231 tgtgttcaaa ttctgtgtct aaggactaca gaacatgtct atagctttcc aaaattgggc taatagtgtg
27301 gctggtcaga atgctattaa agcgtttat gactacacta ccactaaact acctaagatt ggtcagatat
27371 ttggttaagt gttcgctggg attggttaatt taatgattgc ttttgcaaa aacagttcca acatttttga
27441 ttgtttgggt aaattaaact ctcaatttag agcatgggtc gaacaagtag gacaatcaca aggggtttaa
27511 gactttatca gttatgttca agagaatggt cttactatta tgcagttaat cggtaaatatc gtaaaagcat
27581 tagttgcttt tggtagtgc atggctccta tagctagtaa attgttagac tttatcacta atctagctgg
27651 agtatttcaa taaactattc aaacacaccc agctatagca caagttgctg gcgttatggg tatatttagg
27721 ggtgtatttt gggctttaat ggcctcgatt gttgctataa gttagttagt tacaatgtgt tttggtttga
27791 gcttattcag cgtcactgaa aagattttag acttcgtag aacatcaagt ttagttactg gagctacgga
27861 agcatctcaa ttgcgattcg gttcgatttc agcacctatt ttagcagttg ttgcagtaat tggtagcttc
27931 attggtgtcc tcgtttattt atggaaaaca aacgagaact ttagaaatc tatctactgaa gcgtggaaac
28001 ggtttaaacc ggcagtttct ggtgcgattc aaggtgtagc cggctgggtt actgaattgt ggggcaaaat
28071 ccaactcact ttacaaccga taatgcttat attgcaagta ttaggacaaa tattcatgca agttttaggt
28141 gttttgttaa taggcatcat tacaacgctt atgaatatca tacaaggttt gtggacttta attacaattg
28211 cgttccaaag cataggaaca gtgatatccg tagcagttca aatcatagta ggtttgttca ctgctttaat
28281 tcagttgctt acggcgact ctccaggtgc ttgggagact attaaaacta cggttaccaa tgtgcttgat
28351 acgatttggc aatcatgca atcagtttgg gactcaatta tcggcttttt aactggcgta atgaatcgaa
28421 caactttctat gtttgggtaca agttggtcac agatatggag tacaatcact aattttgtta gcagttttg
28491 gaacacgtt gaaagtgggt tcagtcagat ggcttcgagt gtactgaaa aaattgggga agcactaaac
28561 tttattatca caaaaggttc tgaatgggtt tctaactatt ggaatcagat tacaagtttc gcgagtaaa
28631 tagctgatgg gtttaaaaga gttgtctcaa atgtagggtg cggtagtagt gatgcacttg gtaagattaa
28701 aagtttcttc atcagtttct taagtctcgg agcggaatta atcggaagg tagctgaggg tagagccaaa
28771 tctgcgcaca aagtagtcag cgcggttagg gatgcgattt catcagcttg ggactctgta acttcatctg
28841 taagtggaca cggtaggggt agtagcttag gtaagggtt agcgggtatca caagcaaaag taattgctac
28911 agactttggc atgcttctta ataaagact atcctctact ttgacagata gtatagtaaa tccgttaagt
28981 acttctatag acagacacat gactagcgat gttcaacata gcttaaaaga aaataataga cctattgtga
29051 atgtaacgat tagaaatgag ggcgaccttg atttaattaa atcacgcatt gatgacatga acgctataga
29121 ccaactcttc atcagtttct aagggaggtt tggtagttga tagcgacga tatagaagta ataggaaatg
29191 gttcacagta tcgcgtcagt gacaactcct tcaactataa tcaacttgaa gtattgaaat ataacgttac
29261 aggcgcagga tatcatcgta actattctga tatagaggggt attgatggta gatttcataa ttacgctaaa
29331 gaagaactta aaaaagtaga gcttaagata aggtataaag tacctaaaat tgcttatgct tcaactttaa
29401 agtcagacgt ccaagcacta tttgctggac gtttttattt aagggaatta gctacaccag acaattcaat
29471 taagttagag catatattag atataccaaa agacaaacaa gcatttgagc ttgattatgt tgatggacga
29541 caactttttg tagactagt aagtgaaagt atctttgaca caacacaaac atcaggggaa ttttctttgt
29611 cgtttgaaac aacgaaacta ccatattttg aaagtgtcgg ttaatagtag gatcttgaaa gtaataacga
29681 cccgaaaaa tggtaggtac ctgatagatt gcctacaaac gaaggtgata agaggcgta aatgacattt
29751 tacaacacta atcagggaga agtttattat aacggtgatg ttcttttaac acagtttaat cagtttaatg
29821 ttgttgaaat agagtttagt gaagatgtta aagctaatga taaggatgga ttcaacttct atacagataa
29891 aggaaatata tcagttatta aggaagtga tttaaaagcc ggagataaaa taactctcga cggtaaacat
29961 acctatagag gttattttaa tatagattct tttataaaaa ctttagaaca accgggttta tatccaggct
30031 ggaatcgatt caagtctaat aaagttaatga aacaaattac atttagacac aaattatatt ttatagtaag
30101 agtagcctat gccaaattta ttaaaaagtc tacagggtgt agggcacgct attaatgtta gtacaaaagt
30171 aagtaaaaag ctaaaatgaag atagttcttt ggtactaaat attatcgaga acgcgagtag gtttgacgca
30241 ataggtgcta taactaaaat gtggacgac actcatgttg aaggtgaaga tgatttcaac gaatatgtaa
30311 ttgtcatact tgataagctt actattggcg aaaaaataag gcttgatata aagctaggc aaaaagaact
30381 tgatgacctt aacaatttct ggatttacca agagtataac gaaagtttta caggcggtga gttcttcaat
30451 actgtcttta aaggaaacgg ttataagatg gtattacatc caaaagtaga tgcactctaa ttcgagggat
30521 taggcaaaag agatacacga ttagaatatc ttaaaaaagg acttgagcgt tatcatctcg aatatgaata
30591 cgatgcgaag actaaaacgt ttcaatttga tgatgaatta tctaagtttg ccaattatta cattaaagct
30661 ggtgtgaatg ctgataacgt caaaatataa gaagatgcat ctaaatgtta tacctttatt aaaggttatg
30731 gtgattttga tggacaacag acttttgag aagcgggact acaaatgtga ttcaactc ctagcacca
30801 attgataggt aaaagagaag cggccacgct tggtttagga cgtattaaaa aagaagatag tttaaaaaa
30871 gcaatggagt tatgtataaa gaaaagtgtc actgttctta tttctctaga cttttagtag ttacgtgaac
30941 atttccaga agctaaccct aaaatagggt atgttgtag agtgggtgat tctgcatag gatataacga
31011 cttagtgaag atagtcgaaa tcaactacac tagagatgcy tacaataata tcaactaagca agatgtagta
31081 ttaggagact ttcaaaaggc taatcgttat aacaaagcag ttcatgatgc tgcanaat tat gttaaaaagc
31151 taaaactctac aaaatccgac ccatcttaag aactaaaagc attaaacgca aaagttaacg caagtttatc
31221 tataaataat gaattggtta agcagaatga aaaaataaac gctaaagtcg ataagatgaa tactaaaaca
31291 gttacaactg ctaatggtac gatcatgtac gactttacta gtcaatcaag tataagaaac atcaaatcaa

31361 ttggaacgat tggcgactct gtagctagag ggtcgcacgc aaaaactaat ttacacagaaa tgttaggcaa
31431 gaatttgaaa gctaaaaacga ctaattcttgc aagaggtggc gcaacaatgg caacagttcc aataggtaaa
31501 gaagcggtag aaaaacagcat ttatagacaa gcagagcaaa taagaggaga cctaatacata ttacaaggca
31571 ctgatgatga ctgggtacac gggtatttggg caggcgctacc gataggcact gataaaacgg atacaaaac
31641 gttttacggt gccctttgtt ctgcaattga agttattaga aagaataarc cagattcaaa aatactagt
31711 atgacagcta caagacaatg cccatgatgt ggtacaacaa tacgcccgtaa agacacggac aaaaacaaac
31781 taggggttaac acttgaggac tatgtaaacg ctcaaatatt agctttagtg gagttagatg taccagtgtt
31851 tgacgcatac cacacagatt acrttaagcc atacaatcca gcttttagga aagcagagcat ggaggacggc
31921 ttacacccta acgaaaaagg tcacgaggtt attatgtacg agttaatcaa ggattattac agtttttacg
31991 actaaaggag cgaaccaatg gcttacggat taattacaag ttacattca atgacaggtc ggaaaatagt
32061 tgctcaaacat gagtataact atcgcttgtt agatgaaggc atgagcaaac ttgagaaaat gtttatatc
32131 catcaaaaag aagaaatata cgcacactca gcgaaaaaaa ttaataactt gaatgacagt gttgaagatt
32201 atttaacgta tttaaatagc cgttttagca atatgattct agggcataac ggcgacggta tcaatgaagt
32271 aaaagacgcy cgtattgata atacagggtta tggctcataag acattgcaag atcggttcta tcatgattat
32341 tcaacactag atgctttcac taaaagggtt gagaaagctg tagatgaaca ctataaagaa tatcgagcga
32411 cagaatcacg attcgaacca aaccgggaatt tatcactgat ttatcgccat atacaaatgc
32481 agcaatgcga tcattttggg tagacccctag aacgaaaaatt atttatatga cgaagctcgc tccaggtaat
32551 cactacatgt tatctagatt gaagcccaac ggacaattta ttgatagatt gcttgttaaa aacggcggtc
32621 acgggtacaca caatgcgtat agatacattg atggagaatt atggatttat tcagctgtat cggacagttaa
32691 caaaaacaac aagtttgtac gtttccaata tagaaactgga gaaataactt atggtaatga aatgcaagat
32761 gtcagtcgga atataattaa cgacagatat acgtcagcga ttataatcc tatagaaaat ttaatgattt
32831 tcagacgtga atataaagct tctgaaagac aagctaagaa ttcattgaat ttcattgaag taagaagtgc
32901 tgacgatatt tagaaaaggta tagacaaagt attgtatcaa atggatatac ctatggaata cacttcagat
32971 acacaacctt tgcaaggat cacttatgat gcagggtatct tatattggta tacaggtgat tcgaatacag
33041 ccaaccttaa ctacttcaa ggtttcgata taaaaacaaa agaattgtta tttaaacgac gtatcgatat
33111 tggcgggtgt aataataact ttaaaggaga cttccaagaa gctgagggtc tagatatgta ttacgattca
33181 gaaaacaggg gtaaacgact ttaaataggg gtaactattg gacctggtaa taacagacat cactcaattt
33251 attctatcgg ccaaaagggt gtttaaccaat tcttaaaaaa cartgcacct caagtatcga tgaactgattc
33321 aggtggacgt gtttaaccgt taccaataca gaacccagca tarctaagt atattacgga agttggctat
33391 tactatattc atacgcaga gacataaagt gcatttagatt tcccgctacc gaaagcggtt agagatgcag
33461 ggtggttctt ggtatgactg cctggacact ataattggtgc tctaagacaa gtacttacca gaaacagcac
33531 aggttagaatt atgcttaaat tcgaacgtgt cattgacatt ttcaataaga aaaaacaacg agcatggaat
33601 ttctgtccgc aaaaacgcgg ttattgggaa catatcccta agagtattac aaaaattatca gatttaaaaa
33671 tctgtggttt agattttctat atcactactg aagaatacaa acgatttact gattttccca aagactttaa
33741 aggtatttgc ttgtggatat tagaagtaaa atcgaataca ccaggttaaca caacacaagt attaagacgt
33811 aataaacttc cgtctgcaca tcaattttta gttagaact ttgttactgg tggcgttggg aaatggagtt
33881 tattcgaaag aaaggtgggt gaataatgat agtagataat ttttcgaaag acgataaact aatcgagta
33951 caaacaacat cacaatataa tccaattatt gacacaaaca tcagrttcta tgaatcagat agaggaaactg
34021 gtgttttaaa ttgtgcagta actaaagata acagacggtt atctataagt tctgaacatg ttaaaacatc
34091 tatcgtgtta aaaaacgatg attataacgt agatagaggc gcttatattt cagacgaatt aacgatagta
34161 gacgcaatta atggggtgtt gcagtatgtg ataccgaatg aatttttaaa acattcagcg aaggtgcatg
34231 ctacggcatt ctttacacaa aacgggagta ataattgtgt tgttgaactg caatttagct tcaatttga
34301 aaatgattta gttagtgggt ttgatgggtat aacaaagctt gtttatatca aatctattca agatactatc
34371 gaagcagtcg gtaaaagact taaccaatga aagcaagata tggatgatac acaaaacgtta atagcaaaag
34441 tgaatgatag tgcgacaaaa ggcattcaac aaatcgaaat caagcaaaac gaagctatac aagctattac
34511 tgcgacgcaa actagtcaa cacaaagctg tacagctgaa gtcgataaaa tagttgaaaa agagcaagcg
34581 attttgaac gttttaacga agttgaacaa caaatcaatg gcgctgacct tgttaagggt tgttaaaagt
34651 caaattggca aaagtctaaa cttacagatg attacggtaa agcaattgaa tcttatgagc agtccataga
34721 tagcgtttta agcgcaggtta acacatctag gattattcat attactaatg caacagatgc gccagaaaa
34791 acggatatag gcacggttaga gaagccttga caagatgggt ttgatgacgg ttgatgacgg tcttctgttc gatgaatcaa
34861 cttatatact aagcaaatct ggtgtgttag ttgtttatgt tgttgataat aatactgctc gtcgcaacatg
34931 gtacccagac gattcaaacg atgagtacac aaataacaaa atctacggca catggtaccg gttttataaa
35001 aagaatgatg gaaacttaac taagcaattt gttgaagaaa cgtctaacaa cgctttaaat caagctaagc
35071 agtatgtaga tgataaattc ggaacaaacga gctggcaaca acataagatg acagagggga atggctaatc
35141 aattcaagtt aacttaata atgcgcaagg cgaatttggga tatrtaactg ctggttaatta ctatgcaaca
35211 agagtgcggg atttaccagg tagtgttgaa agttatgagg gttatttatc ggtattcggt aaagacgata
35281 caaacaagct atttaacttc acgccttata actctaaaaa gatttacaca cgaatcaatc caaacggcag
35351 acctgagcaa cagtggacag ttccataatga acataagtc aacggattgt tgcagcgtgg agcaaatggg
35421 gtaggtacaa caatcaatct aacogaacca tacacaaact attctatttt attagttaag ggaacttatc
35491 caggtggcgt tattgagggg ttccgactaa ccacattacc taatgcaatt caatttaagt aagcgaatgt
35561 agttgactca gacggtaacg gtggcggtat ttatgagtgt ttactatcca aaacaagtag cactacttta
35631 agaactgata acgatgtgta ctttgattta ggtaaaaacat caggttctgg agcgaatgcc aacaaagtta
35701 ctataactaa aattatgggg tggaaataat gaaactcaca gtaaatgata aaaatgaagt tatcggtatc
35771 gtttaactag gcggtttacg caatagttta gatgtagacg ataacaatgt gtctatcaaa ttcaaaagag
35841 agttcgaacc tagaaagttc gttttcacta acggcgaaat taaatacaat agcaatttcg aaaaagaaga
35911 cgtaccgaat gcatacaacc aacaaagtcg gtcagattta agtgatgagg aacttcggcg aatgggtgca
35981 agtatgcmaa tgcagatgac gcaagtgaac atgttgacaa tgcattgac gcaacaaaac gctatgttaa
36051 cacaacagct gaccgaactg aaaaactaaca aaacaaatac tgagggggac gtttaaatga tgaagatgat
36121 ttactcaact ttaaaagaca ttaaaacttt ttatgtgtgg ggttgcctata aaaaagagca aattcaagtgg
36191 tacgtagaca tgggtgtaat cgcacaaagaa gaatatgcac tgatcactgg tgaataatata ccagaggcaa
36261 aagatgaaaa gtcacaggtg taatgcttga ggctttttaa ttaacacaa agtaggtggc gtaattgttg
36331 gatttaccaa acggcacgaa cartgaatggc gaattagaag attagaagag aatgataaaa caatgcttag
36401 cactctcaat gaggtttaaat taggtcaaaa aactcaagag caagttaaca ttaaataga taaaacttta
36471 gatgctatcc agagggaaag acagatagac gaaaaaataa agaaagaaaa cgcacaaaat atacgcgata
36541 tgaanaatgt gattctcggg ttgataggga ctatcttcag tacgattgtc atagctttac caagaactat
36611 ttttgggtatt taaaggaggt gattaccatg cttaaaggga ttttaggata tagcttctgg gcgtgcttct

36681 ggttttggtaa atgtataata cagtttaagag tcagtgtctt ggcactggct ttttattttg attgaaatga
36751 ggtgcataca tgggattacc taacccaaag actagaagc ctacagctag tgaagtgggt gagtgggcaa
36821 agtcgaatat tggtaagagg ataatatag ataattatcg gggcagtc aa tgggggata cactaaactt
36891 tattttttaa agatattggg gttttgtaac atggggcaat gctaaggata tggctaatta cagatattct
36961 aagggtttcc gattctatcg ttattcatct ggattgtac cggaacctgg agacatcgca gtttggcacc
37031 ctgggcaacgg aatagggttc gacggacaca ccgcaatagt agtaggacca tctaataaaa gttattttta
37101 tagcgttgac caaaactggg ttaattctaa tagttggaca ggttctccag gaagattagt aagacacctt
37171 tatgtaaagt ttacaggctt tgttaggcct ccatactcaa aagatactag caaacctagt agtactgata
37241 caagtccagc atcaaaagcc aatgactcaa caattactgg cgaagcgaag aaaccgcaat ttaagaaggt
37311 taaaacagta aatatactg cttacagcaa tgttttagat aaagaagagc acttcattga tcatatagtt
37381 gtaattgggt atgaacgctc agatattcaa ggattatata taaaagaatc aatgcataag cgttctgtag
37451 acgaactcta tacgcaaga aataagttta taagcgatta tgaataaccg catttatatg tcatatagaga
37521 ggctacatgg cttgctagac caaccaattt tgatgaccgg cgtcacccta attggctagt tattgaagta
37591 tgtgggtggt aacacagatg caaacgacaa tcttattga atcaataca agcgttaata cgtgggtgtt
37661 ggttattgtc agggattgat aaaaacttat ctgaaacgac gttaaaggtt gaccttaata tttggcgtag
37731 tatgaaagat ttaattattt acgacttgat taagcaaggt ataccggata acgcaaaagta tgagcaagtt
37801 aaaaagaaaa tgccttgagac atacattaaa cgagatatat tgacacgaga aaatataaaa gaagtaacga
37871 caaaaacaaac aataagaatt agtgataaaa catcagttga cagtgcgtcc acacgagggc ctactccatc
37941 agacgaaaaa ccaagcatcg ttactgaaac aagtccattc acattccagc aagcactgga tagacaaatg
38011 tctaggggtt acccgaaaaa atctcatata tggggctggg ctaatgcaac acgagcacia acgagctcgg
38081 caatgaatgt taagcgaata tgggaagta acacgcaatg ctatcaaatg ctttaatttag gcaagttaca
38151 aggcatttca gttagtgcgc ttacaaaaat acttaagga aaagggaacgc tgcagcgaca aggcgaagca
38221 ttcgcggaag cttgttaagaa aaacaacatt aacgaaattt attgatcgc gcacgcttcc ttagaaggtg
38291 gatacgggac aagtaacttc gctagtggta gatacgggtc atataattac ttccggtatt gtgcatcg
38361 caacgacctt gattatgcaa tgacgtttgc taaaaataaa ggttggacat ctccagcaaa agcaatcatg
38431 ggcggtgcta tctcgtaag aaaggattac atcaataaag gtcaaaaaac attgtaccga attagatgga
38501 actctaagaa gccagctacc caaccaatcg ctactgctat agagtgtgtc caacatcaag caagtacaat
38571 cgctaagtta tctaacaaca tcgggttaaa aggtatctac ttccacaggg ataaataaa ataaagaggt
38641 ggttaaatgt acaaaataaa agatgtgaa acgagaataa aaatgatgg tgttgactta ggtgacattg
38711 gctgtcgatt ttacactgaa gatgaaaata cagcatctat aagaataggt atcaatgaca aacagggtcg
38781 tatcgatcta aaagcacatg gcttaacacc tagattacat ttgtttatgg aagatggctc tatattcaaa
38851 aatgagccccc ttattatcga cgatgttga aaagggttcc ttacctacaa aatacctaaa aaggttatca
38921 aacacgctgg ttatgttcgc tgaagctgt ttttagagaa agaagaagaa aaatacatg tcgcaaatct
38991 tcttttcaat atcgttgata gtggatttga atctgctgta gcaaaagaaa tcatgtttaa attggtagat
39061 gatgctatc cgagaatttt aaagataaac cgcacagatc tattgagcaa agactttaaa gagaaatag
39131 ataagatgtt cttctcttac atcgaaaaaga atgaagtag atttaagggt gcgaagaggt ataaaggcga
39201 accgggacaa cctggtgcga aaggtgatac aggtaaaaaa gggaacaaag gcgcaccagg taaaaacggt
39271 atcgttagat caatcaatcc tgacactaaa atgtggcaaa ttgatggtaa agatcacgat atcaagcag
39341 aactctgatt attggaacaa atcaatatcg caaatgttga agggtagaa gataaattgc aagaagttaa
39411 aaaaatcaaa gatcacactc tcaacgactc taaaacgat acggattcaa aaattgtcga actagttgat
39481 agcgcgcctg aatctatgaa tacattagaa gaattagcag aagcaatata aaacaactct atttcagaaa
39551 gtgtattgca acagattggc tcaaaagtta gtacagaaga ttttgaggaa ttcaaaacaa cactaaacga
39621 tttatattgct ccaaaaaatc ataactatga tgacgggtat gttttgtcat ctcaagcttt tactaaacaa
39691 caagcggata atttatatca actaaaagc ccatctcaac cgacgggtta aatttggaca ggaacagaaa
39761 atgaatataa ctatatatat caaaaagacc ctaatacact ttacttaatt aaggggtgat ttttatggaa
39831 ggttaatttta aaatgtataa gaagtttatt tacgaaggtg aagaatatac aaagtatat gctggaaata
39901 tccaagatc gaaaaagcct tcatcttttg taataaaacc cttaactaaa aataaatatc cggatagcat
39971 agaagaatca acagcaaaat ggacaataaa tggagttgaa cctaataaaa gttatcaggt gacaatagaa
40041 aatgtacgta gcggtataat gaggttttcg caaactaatt taggttcaag tgatttagga atatcaggag
40111 tcaatagctg agttgcaagt aaaaatatca acttttagta tcttccagg atgttgtatg tcaataaag
40181 tgatgtttat tcaggatctc caacattgac cattgaataa ttttaaacga ctaatttttt agtcgttttt
40251 tattttggat aaaggagca aacaaatgga tgcaaaagta atacaagat acatcgattt gatcttagca
40321 ttatgtaaat aattcttagc gaacaaaggt attagccga ttccagtaga cgatgagact atatcatcaa
40391 taatacttac tgttgttgct ttataacta cgtataaaga caatccaaca tctcaagaag gtaaatgggc
40461 aatcaaaag ctaaaagaa ataaagctga aaacaagtat agaaaagcaa caggggcaagc gccaattaaa
40531 gaagtaatga accctacgaa tatgaacgac acaaatgatt tagggtaggt gttgaccaat gttgataaca
40601 aaaaaccag cagaaaaatg gtttgataat tcattagggga agcagttcaa tctgtatttg ttttatggat
40671 ttcatgttta cgtattcgca aatatgtttt ttatgatagc aacaggcgaa aggttacaag gtttatacgc
40741 ttataatatt ccatttgata ataaagcaag gattgaaaaa tacgggcaaa taattaaaaa ctatgatagc
40811 tttttaccgc aaagtggga tattgtcgtt ttcctgtcaa agtatgttg cggagctgga catgttgaaa
40881 ttgttgagag cgcataatca aacactttca catcatatgg gcaaaattgg aatggtaaa gttggacaaa
40951 tggcgttgag caactgggtt ggggtcctga aactgttaca agacatgttc attattacga tgaccaaatg
41021 tattttatta gattaaattt cccagataaa gtaagtgttg gagataaagc taaaagcgtt attaagcaag
41091 caactgccaa aaagcaagca gtaattaaac ctaaaaaaat tatgcttgta gccggctatg gttataacga
41161 tctctggagca gttaggaacg gaacaaacga acgctgattt atccgtaaat atataacgcc aaatctcgt
41231 aagtatttaa gacatgcagg tcatgaagtt gcattatatg tgggtcaaag tcaatcacia gacatgtatc
41301 aagatattgc atacggtgtt aatgtaggaa ataataaaga ttatggatta tattgggtta aatcacaggg
41371 gtatgacatt gttctagaga tcatattaga cgcagcagga gaaaatgcaa gttgtgggca tgtttatttc
41441 tcaagtcaat tcaatgcgga tactattgat aaaagtatac aagatgttat taaaaatac ttaggacaaa
41511 taagaggtgt aacacctcgt aatgatttac tgaacgttaa tgtatcagca gaaataaata tcaattatcg
41581 tttactgaa ttacttttta ttactaataa aaaagatag gattggatta agaagaatta tgacttgtat
41651 tctaaattaa tagctgggtc gattcatggt aagcctatag tgggttgggt agctggtaat gttaaaacat
41721 cagctaaaaa ccaaaaaat ccaccagtg cagcaggtta tacacttgat aagaataatg tgccttataa
41791 aaaaagagct gctaattaca cagttgccaa tgttaagggt aataacgtaa gggacggcta tccaactaat
41861 tcaagaatta caggtgtatt acctaatac gcaacaatca aatatgacgg cgcattatgc atcaatgggt
41931 atagatgatt tacttatatt gcaaataggt gacaacgtcg ctatattgcy acaggagaggt tagataaagc

42001 aggttaatagg ataagtagtt ttggttaagtt tagcacgatt tagtatttac ttagaataaa aattttgcta
42071 catataattat agggaaatctt acagttatta aataactatt tggatggatg ttaatatccc tatacacttt
42141 ttaacattac tctcaagatt taaatgtaga taacaggcag gtactacggt acctgcctat ttttttgta
42211 taatgtaatt acattaccag taaccaatct ggctttaaacc cacatttccg gttagccaatc cggctatgca
42281 gaggacttac ttgcgtaaag tagtaagaag ctgactgcat atttaaacca cccatactag ttgctgggtg
42351 gttgtttttt atgttatatt ataaatgac aaaccacacc acctattaat ctaggagtgt ggttattttt
42421 tatgcaaaaa aaacgaaaaa aagttcataa aaagtattgc atatcacgtt taaccgtgtt ataataaggt
42491 ataccagtgt agaggaggat aaaaagtgtt agaaaaattt aaaaactatag cagaaatcgc cttttatata
42561 atgtcagcaa ttgccatagc gaaaacattg aaaaaagacg ataagtaagt agacaagccc gaaagggctg
42631 tctatatata aattctaaca ctaaaatact atgaaaaaaa tttacattat tttaatcatt cttatttgga
42701 taaaagtggt tttaggcaac gatataagta aaagtgtgtt tgcactgctt actactttac tgcttatcaa
42771 tttatggaag agggataaaa atgacagcaa taaaagaaat aattgaaatca atagaaaagt tattcgaaa
42841 agaaacggga tataaaattg ctaaaaattc cggattacca tatcaactg tgcaagattt aagaaatgga
42911 aaaacatctt tatcagatgc cagattttaga acgataataa agttatacga gtatcaaaaga tcgcttgaaa
42981 acgaagaaga taataaaaag gagccaaaaa tatgtttgtt acaaaagaag aatttaaaac tttgaatgta
43051 aaagaagtat ttgaatcagg taaaaacttt ataaaaatta cagatggaag acatgcaata tattgggtaa
43121 atgatagata cgtagtactt gaccataaaa aaggcgtatt gtaccgcaa aaagcatacc caaaatata
43191 caaaagaaaa ttagtaagtt aaataattag aaaaaccagt cttaatgac gtggttattt ttaggtttg
43261 cgcgtgtcaa atacgtgtca atttagttct atttcttttag ttttctttct aaacttaatt gcttgtaaac
43331 cgcatagtta taggcttttc agctatatac caagataaga tttatccgc cgtctccata aaaatatgct
43401 tggaaacctt gatttaattg ggttttaatc tagcaagtgt caaatatgtg tcaagaaaat aattttctga
43471 cacgttgacc ttgctctttt ttatgttcat caagtaagt agagtaggtg tctaaagtta tagatatatt
43541 ataattggctt aatcttttgc taatatattc aatagg

Table 10

Bacteriophage 96 ORFs list

SID	LAN	PRA	POS	a.a.	RBS sequence	STA	STO
100733	96ORF001	1	25999..29142	1047	ccctgaatcgaaaggaggttagcct	ttg	taa
100734	96ORF002	1	32008..33906	632	tttttacgactaaaggaggaacca	atg	taa
100735	96ORF003	1	30109..31995	628	ttatatttttagataaaggagttagcct	atg	taa
100736	96ORF004	1	36760..38634	624	attttgattgaatgaggtgcatac	atg	taa
100737	96ORF005	3	33903..35729	608	gtttattcgaaaggaaaggtggttga	ata	taa
100738	96ORF006	2	40589..42043	484	aatgattttagggttaggtcttgacca	atg	tag
100739	96ORF007	1	18652..20091	479	tatacacacatactaaacctgaacg	att	tga
100740	96ORF008	2	8960..10201	413	tggcagaatttggggcgataaacga	atg	tga
100741	96ORF009	2	17447..18670	407	gacgcaataaacggaagtcatcgtca	atg	tga
100742	96ORF010	1	38647..39819	390	taaatataaataaaggaggtgtgtaa	atg	tga
100743	96ORF011	-1	119..1195	358	gtagctcgccctacccttattatttt	ttg	tga
100744	96ORF012	2	20045..21013	322	tttaatgacaaattacctgacatag	atg	tga
100745	96ORF013	3	29157..30098	313	acttattataaggaggtttgttag	ttg	taa
100746	96ORF014	1	21925..22839	304	agaaaaataaagtgaaggtataaata	atg	tag
100747	96ORF015	1	5812..6591	259	atacacggtaaaagggtgggagaaatag	atg	taa
100748	96ORF016	1	7852..8607	251	aataaaatgttgaaaggagagaaaa	atg	taa
100749	96ORF017	3	3444..4190	248	aaatttaacatttaatatcactttaa	gtg	taa
100750	96ORF018	-3	28281..29000	239	taagctatgttgaaacatcgctagtc	atg	tga
100751	96ORF019	3	7188..7859	223	tttaccggtctaggacgtgtgttaa	atg	taa
100752	96ORF020	3	21324..21908	194	gaaggggcaaaaaggagttttgatat	atg	taa
100753	96ORF021	3	6612..7175	187	attaaaaatttaattaaaaggacggt	ata	tag
100754	96ORF022	2	24536..25093	185	aaagaaaaacgaaggagtgatttaa	atg	taa
100755	96ORF023	1	5275..5811	178	catgaaatggttaggaggtatgaaaa	gtg	tag
100756	96ORF024	3	14481..15014	177	taaaacgtaggagataaacgaataa	atg	taa
100757	96ORF025	2	25157..25666	169	ataaaaaaattgaaagagggtatat	att	taa
100758	96ORF026	-3	15084..15590	168	tcattcttaacatagcccttaattc	atg	tga
100759	96ORF027	-1	1229..1732	167	aatagcaataaaggagtgtaaaac	atg	taa
100760	96ORF028	1	16960..17454	164	aaaggcgtgtgatacagtgaaacaa	ttg	taa
100761	96ORF029	-1	1736..2227	163	tatgagaaaaaggagtcataataaaag	atg	taa
100762	96ORF030	1	25531..25995	154	ttttcaaggaggagagtcgctcgtta	ctg	tag
100763	96ORF031	2	23633..24097	154	tttagtattgaagggtgattctgttag	atc	tag
100764	96ORF032	-2	2248..2706	152	ataagacaccaaaagggttttggcgc	atg	tga
100765	96ORF033	-3	39147..39605	152	agcatataaattcgtttagtgtttgt	ttg	taa
100766	96ORF034	2	13181..13615	144	tagaagtcgaaaaagtgaggcaat	ata	taa
100767	96ORF035	2	10628..11053	141	gagctaggatttgcagcaacgatat	ttg	tga
100768	96ORF036	2	24110..24535	141	gtatttttcatagaggtggttaaat	atg	taa
100769	96ORF037	1	12583..12996	137	atgaggaacagaaagcaaccaacttt	att	tga
100770	96ORF038	1	15628..16032	134	atgttaagaatgatgcctagttttaa	ttg	taa
100771	96ORF039	3	39816..40220	134	ctaatacactttacttaattaaagg	gtg	taa
100772	96ORF040	-3	27528..27932	134	tttccataaataaacgaggacacca	atg	tga
100773	96ORF041	3	16206..16607	133	gatgagggcgagggtgtcagagtag	atg	tga
100774	96ORF042	2	35720..36106	128	aagttactataactaaaattatggg	gtg	taa
100775	96ORF043	-2	35713..36081	122	ttaaacgtccccctcagttattgtt	ttg	taa
100776	96ORF044	-2	9460..9828	122	agtatccatcagttgaagataatct	ata	taa
100777	96ORF045	-3	5139..5504	121	ctctttttgtattctgtaatttca	att	tga
100778	96ORF046	2	11513..11872	119	aagtaaatgtatagaggtggaataa	atg	taa
100779	96ORF047	2	22991..23350	119	gtcgtactacgtctgataaagagcga	gtg	tag
100780	96ORF048	3	8607..8963	118	tggaaaaagaatttgagtgatgacta	atg	tga
100781	96ORF049	1	23353..23697	114	atccggttaaaccaataaggtagag	gtg	taa
100782	96ORF050	-2	2728..3072	114	tggttaattagttattacattaaagta	ata	taa
100783	96ORF051	3	4692..5021	109	tcaaaatatacggaggttagtcaact	atg	tga
100784	96ORF052	-1	20882..21211	109	gtagcaaaagagacaactaaaaaagt	gtg	taa
100785	96ORF053	1	40252..40578	108	acgactaatttttttagtcgtttttt	att	tag
100786	96ORF054	1	4942..5262	106	aatataaaactaaaaacaaaattt	atg	tag
100787	96ORF055	-2	4840..5151	103	ccgtcccaatatatagttccgttaa	atc	taa
100788	96ORF056	3	36324..36623	99	aatttaacacaaagtaggtggcgta	atg	taa
100789	96ORF057	2	1394..1690	98	cttcagtggtctcttttagcatttaa	ata	taa
100790	96ORF058	-3	26247..26537	96	tacttttttttccataactcagacca	att	tga
100791	96ORF059	-1	21485..21772	95	agactcaacgcctttttgaaacatac	ttg	tga
100792	96ORF060	-3	22647..22931	94	cctcrttgttaaccgacaagactgta	ata	taa
100793	96ORF061	1	14023..14304	93	ttatctaataaaggggacgagtaga	gtg	taa
100794	96ORF062	-2	38281..38559	92	tatacaacttagcgattgttacttgc	ttg	taa

100795	96ORF063	-3	30786..31064	92	gtctcctaactactacatcttgctta	gtg	tga
100796	96ORF064	-2	30205..30480	91	atgcattctacttttggatgtaaac	ata	tag
100797	96ORF065	1	2617..2886	89	aaggtcttaataaaaaatttctccttc	ttg	tga
100798	96ORF066	3	28056..28325	89	aaggtctagtcggctgggttaactga	att	tga
100799	96ORF067	-3	17142..17411	89	ttccgttattgctgctggaagttgt	ttg	tga
100800	96ORF068	2	12326..12589	87	aatgcargtcgtttggtctgcctaa	ttg	tag
100801	96ORF069	2	42734..42997	87	tttttaggcaacgataataagcaaaa	gtg	tga
100802	96ORF070	1	11869..12129	86	aaatgttcaagaaatggagtgaaagc	ata	tga
100803	96ORF071	3	15396..15656	86	aacaagctatacaaaattatcgataa	att	tga
100804	96ORF072	-3	37749..38009	86	agattttttcgggttacccttagac	att	tga
100805	96ORF073	3	11244..11501	85	acatgcatacatagagggtggaataa	atg	tag
100806	96ORF074	-3	42936..43193	85	aattatttaacttactaattttctt	ttg	tga
100807	96ORF075	-3	26610..26867	85	tactgccaatgttccatcttcaacc	att	tga
100808	96ORF076	-1	11126..11380	84	tttatctaatacatttaagttaacc	atc	tga
100809	96ORF077	-2	16537..16791	84	taccaccatataaggcaggtagtag	gtg	tag
100810	96ORF078	-3	19521..19775	84	aataactttgaatttgatcctcaac	ata	tga
100811	96ORF079	3	13608..13859	83	ttagggaatggaggcagacacaa	atg	tag
100812	96ORF080	-3	28029..28280	83	tgagaagtcgccaagtaagcaactga	att	tga
100813	96ORF081	3	20973..21221	82	aatgaagttatcccatctcatgactt	atc	tag
100814	96ORF082	-1	8729..8974	81	cgattattgtgctttcaatttcaaa	ttg	tga
100815	96ORF083	-3	3147..3392	81	tttagcctttatataatcaactctc	gtg	tga
100816	96ORF084	3	1611..1853	80	tgctttatcttttagttttcttctt	ttg	tga
100817	96ORF085	-2	29470..29709	79	ctctttatcaccttcgtttgtaggca	atc	tga
100818	96ORF086	1	35188..35424	78	gcgcaaggcgattttgggatatttaa	ctg	tag
100819	96ORF087	-2	13039..13275	78	ttttgattgagctctaaagtgtctt	att	tag
100820	96ORF088	3	24930..25163	77	gaactatcattaaaagttaaatgga	ata	tga
100821	96ORF089	-3	22329..22562	77	tccagttataagatagtggttaatccc	ata	tga
100822	96ORF090	-3	16803..17036	77	accttttagtcgaataaccctgcgtca	ata	tag
100823	96ORF091	-1	22559..22789	76	aacgcttctgtgtttaacgttcatgt	atg	tga
100824	96ORF092	3	18360..18587	75	attgcaaaagatattgtaagttagat	atg	tga
100825	96ORF093	-2	25384..25608	74	catgatttccttgtattctctcttc	atc	tga
100826	96ORF094	1	10417..10638	73	aacacacattaaaggagtggttaaaaa	atg	tag
100827	96ORF095	3	12963..13184	73	tactaaacgaagataaaaactatgac	att	tga
100828	96ORF096	1	42994..43212	72	gatcgcttgaanaacgaagaagataa	ata	tga
100829	96ORF097	-1	36047..36265	72	tcaagcattacacctgtgacttttc	atc	tga
100830	96ORF098	-2	36766..36984	72	cagggtccgggtacaaatccagatga	ata	tga
100831	96ORF099	-2	34765..34983	72	ctattctttttataaaacgggtacc	atg	tag
100832	96ORF100	1	10198..10413	71	acaagaagactcagaggtttttcac	atg	tga
100833	96ORF101	1	15208..15423	71	gagaanaacagtttaagataaggagag	atg	tga
100834	96ORF102	3	4209..4424	71	attttaaaacgaataataggagagg	ctg	tag
100835	96ORF103	3	11673..11888	71	catgcaccttatgggtatgcgcttag	ctg	tga
100836	96ORF104	3	12117..12332	71	ttttagctccaaagagcttttgactt	gtg	tga
100837	96ORF105	3	23892..24107	71	gatggtgggttatccagtggtataa	gtg	tga
100838	96ORF106	-3	34428..34643	71	tagacctttgccaattttgtgttga	att	tga
100839	96ORF107	-3	24495..24710	71	ggcacattaccaattgttaatttaa	atg	tga
100840	96ORF108	-1	23876..24088	70	acatatttaaccacctctatgaaaa	ata	tga
100841	96ORF109	-2	17317..17529	70	acctgtacgcttttgcctccgtgatta	att	tga
100842	96ORF110	-3	38931..39143	70	accttcattcttttcgatgtaagaa	atg	tga
100843	96ORF111	-3	21855..22067	70	agtaaaatttttcttttgcgtgtc	att	tga
100844	96ORF112	1	3217..3426	69	aaatgtcaacgggaggtgatacga	atg	tga
100845	96ORF113	-1	25469..25678	69	tcagggtatatactctaaatatctag	ctg	tga
100846	96ORF114	-2	9838..10047	69	ataataatcatcacggtaaaagttagc	atc	tga
100847	96ORF115	1	13819..14022	67	gcagtaggggttatggcagggtcaag	ttg	tga
100848	96ORF116	-1	41033..41236	67	caacttcatgacctgcgtgtcttaa	ata	tga
100849	96ORF117	-3	24711..24914	67	tctgctgtattccatttaactttta	atg	tga
100850	96ORF118	-1	12374..12574	66	tccatctctcttaaaataaagtgtg	ttg	tga
100851	96ORF119	-1	3980..4180	66	ctcctatatttcgtttttaaatttc	att	tga
100852	96ORF120	-3	6033..6233	66	ttgtaatttagaaatataaacgataa	ata	tga
100853	96ORF121	-2	37939..38136	65	ctgaaatgccttgatacttgcttaa	att	tga
100854	96ORF122	2	37892..38086	64	acgacaaaaacaacaataagaatta	gtg	tga
100855	96ORF123	-3	29193..29387	64	ggacgctgacttttaaatgtgaagc	ata	tga
100856	96ORF124	1	4408..4599	63	tttatcggtaccaatttaaatgatta	atg	tga
100857	96ORF125	-1	7787..7978	63	ttaaaaatccaagttttgccatcgt	att	tga
100858	96ORF126	-3	27027..27218	63	aaatttgaacaacggcatttaattga	gtg	tga
100859	96ORF127	3	15051..15239	62	atcgagtcgaaggaggttttggggaa	gtg	tga
100860	96ORF128	-1	6914..7102	62	agcgaatgggttatgattgtgactc	ata	tga
100861	96ORF129	-3	31332..31520	62	tcttatttgcctctgctgtctataa	atg	tga
100862	96ORF130	-3	30084..30272	62	gaaatcatcttcaccttcaacatga	gtg	tga
100863	96ORF131	3	11058..11243	61	agaaaaagagaatgaagtgtatcta	atg	tga
100864	96ORF132	-1	36434..36619	61	taagcatggtaaatcacctccttaa	ata	tga
100865	96ORF133	-1	35591..35776	61	ctaaactattgcgtaaacccgctcta	att	tga
100866	96ORF134	-2	9250..9435	61	atccatgagcttataaccgctcta	att	tga

100867	96ORF135	1	29563..29745	60	cgacaactcttttctgtaggactagtaa	gtg	tga
100868	96ORF136	-3	12486..12668	60	cactttacttttcaactctgttcaggga	ttg	taa
100869	96ORF137	-1	14501..14680	59	caaaactgaaagcctaagtaatacagca	atc	tga
100870	96ORF138	-2	23326..23505	59	cttctgacatttctgatgaaatttttag	ttg	tga
100871	96ORF139	-3	42672..42851	59	aaacccggaatttttagcaatttttat	atc	taa
100872	96ORF140	-3	31137..31316	59	acttgattgacttagtaaaagtcgtac	atg	taa
100873	96ORF141	-3	18969..19148	59	aacaaaaataacattatagggatct	ata	taa
100874	96ORF142	-3	4740..4919	59	cataaattttgttttttagtttttat	att	tga
100875	96ORF143	2	36107..36283	58	aacaaatactgagggggacgttttaa	atg	taa
100876	96ORF144	3	16029..16205	58	tatcacgaagtaaaagaaggtagataa	ata	tag
100877	96ORF145	-3	29013..29189	58	tgctcactgacgcgataactgtgaacc	att	tga
100878	96ORF146	-3	14883..15059	58	aatcttctgaattgttgactaagta	ttg	taa
100879	96ORF147	-1	18251..18424	57	tatcacgcgttaattgcacgtaattct	atg	taa
100880	96ORF148	-1	13583..13756	57	aataccttctttaactgaattgttga	ata	taa
100881	96ORF149	-2	10756..10929	57	taaattccacatctctatactgatat	ctg	tag
100882	96ORF150	2	14171..14341	56	atttttaagtgaagaagtgtttattaa	ctg	tag
100883	96ORF151	2	19217..19387	56	cctacatactcattgcgctactttt	atg	tga
100884	96ORF152	-1	12614..12784	56	attttctacagtaaaaaatctcttat	ctg	taa
100885	96ORF153	-2	11936..12006	56	ttgcattacctattgcgaattgctag	ttg	taa
100886	96ORF154	-2	4165..4335	56	atataacgctttttctcctcgaccaa	atc	tga
100887	96ORF155	-3	40464..40634	56	aaatcaggattgaaactgcttcccta	atg	tga
100888	96ORF156	3	423..590	55	tggttaattttgataatttagcttta	ata	taa
100889	96ORF157	-1	41879..42046	55	gtagcaaaatttttattctaaagtaa	ata	taa
100890	96ORF158	-2	36166..36333	55	cattcatgttcgtgcccgtttggtaa	atc	tag
100891	96ORF159	-2	16228..16395	55	tttaacatctgagcataccttttat	ttg	taa
100892	96ORF160	3	1038..1202	54	atctctaagcagttgtrtgagcagcg	ttg	taa
100893	96ORF161	-1	19193..19357	54	tcctttgtgttaggtacaccaaaca	atg	tag
100894	96ORF162	-1	18074..18238	54	cccgctcctattaacacataagatcc	ata	tga
100895	96ORF163	-1	15386..15550	54	agccatcataggactgtaaaattca	ctg	taa
100896	96ORF164	-1	10049..10213	54	tacatcgatttcaataagcttttga	att	tag
100897	96ORF165	-2	18514..18678	54	gtgcttcaatatcatctattaaact	ata	taa
100898	96ORF166	-2	11104..11268	54	ctagccatgattacccttaaatag	ttg	tag
100899	96ORF167	-3	13764..13928	54	agacagtttataatgtgtatctcta	ata	tga
100900	96ORF168	1	14305..14466	53	ttttgaatttttggaggacgagtaa	atg	tag
100901	96ORF169	-1	17885..18046	53	gtggtgaagccttaataagactctt	ata	tga
100902	96ORF170	-1	10790..10951	53	taggcgctttacatattccacgttaa	att	taa
100903	96ORF171	-3	12765..12926	53	atcttcgcttttagtatataaaacgct	ctg	taa
100904	96ORF172	3	22836..22994	52	cgctcgcaacgcttaaaccaactga	ata	tga
100905	96ORF173	-1	15956..16114	52	ctctacatcatcattagccgtctgc	ata	taa
100906	96ORF174	-1	10571..10729	52	tagtgccattcatattactttcttaa	ata	taa
100907	96ORF175	-1	3440..3598	52	cagcctatcttcaactatcaacatga	ttg	taa
100908	96ORF176	-3	37170..37328	52	tttatctaaaacattgctgtgaagca	gtg	taa
100909	96ORF177	-3	6693..6851	52	ttcctaattctactaagtaactcgat	ata	taa
100910	96ORF178	-3	5655..5813	52	gacatcttgattagttttttcagtc	atc	tag
100911	96ORF179	1	34564..34719	51	gttcacagctgaagtcgataaaaatag	ttg	tag
100912	96ORF180	1	42661..42816	51	atataaattctaacactaaaatatt	atg	tga
100913	96ORF181	-2	37741..37896	51	tggaacgactgtcaactgatgtttt	atc	taa
100914	96ORF182	-2	25039..25194	51	ttcgtaattctttttctccgtcatta	atc	tga
100915	96ORF183	-2	4534..4689	51	tcagtttttaatttttcagccatag	ttg	tga
100916	96ORF184	1	6721..6873	50	ggagctgggaatttacagtaaaaag	ttg	tag
100917	96ORF185	2	36548..36700	50	acaaaaataacgcgatatgaaaaat	gtg	taa
100918	96ORF186	-1	40025..40177	50	tggaagatctgaataaacatcactt	ata	tga
100919	96ORF187	-1	34466..34618	50	attacctttaacaaggtcagcgcca	ttg	tga
100920	96ORF188	-1	33842..33994	50	agttcctctatctgattcatagaaa	ctg	taa
100921	96ORF189	-1	24914..25066	50	acatagaatgggtcttccgtgtgtga	atc	taa
100922	96ORF190	-2	20395..20547	50	tatcttagagtaaacctctccactc	ata	tga
100923	96ORF191	3	24768..24917	49	aaaggaattgaagcagtgaaacacg	ctg	taa
100924	96ORF192	-1	16169..16318	49	tcgtgggttcggcaacgtttgcttgt	atg	tga
100925	96ORF193	-2	39100..39249	49	cagtaccgtttttaccgggtgcgc	ttg	taa
100926	96ORF194	-2	25921..26070	49	ttggtacagaogtctttgctaatcg	ttg	taa
100927	96ORF195	-2	17779..17928	49	caaccaatgctcgggatggtcaggg	ttg	tga
100928	96ORF196	-2	14182..14331	49	ttaaaatacttttcttctagcaatgc	atc	tga
100929	96ORF197	-2	7609..7758	49	ttatcatcaaacgacttaacaccaa	ttg	tga
100930	96ORF198	-2	1537..1686	49	ttattagctagtgcgttagtgttag	gtg	taa
100931	96ORF199	-3	7719..7868	49	caatacttgtatcggatagtcattct	att	taa
100932	96ORF200	2	22271..22417	48	ttctttaatgaggttaaacctcttaa	ttg	tga
100933	96ORF201	2	30353..30499	48	cttactatttggcgaaaaataagcg	ttg	tag
100934	96ORF202	2	32591..32737	48	agattgaagcccaacggacaaattta	ttg	taa
100935	96ORF203	2	39131..39277	48	agcaaaagactttaaaagaaaaatag	ata	tag
100936	96ORF204	-2	36985..37131	48	atcttccctggagaaacctgtccaa	atc	tga
100937	96ORF205	-3	38721..38867	48	aaggaacctttttacaacatcgtcg	ata	taa
100938	96ORF206	-3	35880..36026	48	gttaacatagcgttttcttgcgtca	att	taa

100939	96ORF207	-3	11550..11696	48	ttgctctctcgtccatgattttgg	ata	taa
100940	96ORF208	2	37178..37321	47	agattagtaagacacccttatgtaa	gtg	taa
100941	96ORF209	2	42341..42484	47	tgcataatttaaacaccaccatactag	ttg	taa
100942	96ORF210	3	41850..41993	47	aaaggtaataacgtaagggacggct	att	tag
100943	96ORF211	-1	6662..6805	47	ttgttggaaatgggtggacgaattgg	ttg	tga
100944	96ORF212	-2	25213..25356	47	agtagcacattcccaaaattgtaaa	atc	taa
100945	96ORF213	-3	42219..42362	47	gtggtttgatcattttataataaac	ata	taa
100946	96ORF214	3	27834..27974	46	aaaagatttttagacttcgttagaac	atc	tag
100947	96ORF215	3	35811..35951	46	ttacgcaatagtttagatgtagacg	ata	taa
100948	96ORF216	-1	5402..5542	46	tttccgtaagggtgattccaacttga	att	tga
100949	96ORF217	-2	24229..24369	46	cataggctctgttaagcacataacct	atc	taa
100950	96ORF218	-2	6253..6393	46	ttgtcattcttgctaacacgtcaga	ttg	taa
100951	96ORF219	1	883..1020	45	aaatcactcccgaaatattcgtaa	ata	taa
100952	96ORF220	2	32936..33073	45	gataaaaggtatagacaagatttctg	atc	taa
100953	96ORF221	3	41703..41840	45	ggtaagcctatagggtggtttggtag	ctg	taa
100954	96ORF222	-1	39860..39997	45	acttttattagggttcaactccattt	att	taa
100955	96ORF223	-1	24716..24853	45	acattttcaaatgattcttggaacac	ata	taa
100956	96ORF224	-2	26794..26931	45	caatatcacgccatgtagttttttaa	ctg	taa
100957	96ORF225	-2	19201..19338	45	caacaatggattgtaatcaaataa	atg	tga
100958	96ORF226	-2	15709..15846	45	tgacttgccttgccttcaacacaaat	ata	taa
100959	96ORF227	-3	36711..36848	45	acattgactgccccgataaattatct	ata	tga
100960	96ORF228	3	2325..2459	44	tcgccatagtaggttccaataaccgt	ata	taa
100961	96ORF229	-1	38612..38746	44	ttgtcattgatacctattcttatag	atg	tga
100962	96ORF230	-1	31733..31867	44	gctggattgtatgcgttaaaagtaat	ctg	tag
100963	96ORF231	-2	12076..12210	44	tgacctatagctttaacttgcctgt	ctg	taa
100964	96ORF232	-3	31644..31778	44	atagtcctcaagtggttaaccctagt	ttg	taa
100965	96ORF233	-3	23988..24122	44	atttgatttgaagtgcaggtcaca	ctg	taa
100966	96ORF234	-3	17529..17663	44	agtagcttttttgaatcgtaaccta	atg	taa
100967	96ORF235	1	7153..7284	43	aatgctaattgggtccaatagaaatca	atg	tag
100968	96ORF236	2	2681..2812	43	ttctttcacttcaacttcacatttc	ata	tga
100969	96ORF237	2	4496..4627	43	gtacctatgcttcacagctcttagcga	ttg	taa
100970	96ORF238	-1	41720..41851	43	cacctgtaattcttgaattagttga	ata	tga
100971	96ORF239	-1	35324..35455	43	acttactaataaaatagaatagttt	gtg	taa
100972	96ORF240	-1	8570..8701	43	atccccggttttgacttaataacatca	atc	tga
100973	96ORF241	-2	33502..33633	43	ataattttgtaataactcttagggat	atg	tag
100974	96ORF242	-2	23662..23793	43	agctaattgctacagcaggtgtgttaa	atc	tag
100975	96ORF243	-3	32391..32522	43	acctggacgagcttgcgcataataa	ata	tag
100976	96ORF244	-3	30273..30404	43	aaaactttcgtttataactcttggtaa	atc	tga
100977	96ORF245	-3	5895..6026	43	tgcaactaaaatgcttataattctcta	atc	taa
100978	96ORF246	-3	2679..2810	43	attcatcaagaaactatagccggtc	atg	tga
100979	96ORF247	1	34891..35019	42	acatcaagcaaatctggtgtgttag	ttg	taa
100980	96ORF248	2	30668..30796	42	aattattacattaaagctggtgtga	atg	tag
100981	96ORF249	2	31838..31966	42	caaatattagcttgtagtgagttag	atg	taa
100982	96ORF250	2	33539..33667	42	cttaccgaacacagcaggtagaa	ata	taa
100983	96ORF251	-1	20486..20614	42	cttctgtacgagccacacgcaatga	ttg	cag
100984	96ORF252	-1	15128..15256	42	gatatttcattacttagctactacta	ata	tga
100985	96ORF253	-2	41446..41574	42	aaaaccttaattcagataaaacgataa	ttg	tga
100986	96ORF254	-2	41005..41133	42	gttataacctatgaccggtacaagc	ata	taa
100987	96ORF255	-2	23008..23136	42	aggataaatgacttgaccatcttctc	ata	taa
100988	96ORF256	-2	14794..14922	42	ttgtatgcgtcaatgagttggtcga	ttg	tag
100989	96ORF257	-2	8503..8631	42	tacctaaacttttttaataattttcta	atg	tga
100990	96ORF258	-3	22143..22271	42	aaacgctttgttaaaatgcctctgca	att	tga
100991	96ORF259	-3	18639..18767	42	cttgratctattatagagatttaacc	att	tag
100992	96ORF260	-3	15624..15752	42	gttttcgtaactagccactgtatag	ata	taa
100993	96ORF261	2	18746..18871	41	catattgaggtcttaatatagatcac	ata	taa
100994	96ORF262	-1	13067..13192	41	aatttaattattcttctcttggttgg	ttg	taa
100995	96ORF263	-2	18742..18867	41	taacagacagctctaatgccttac	att	tga
100996	96ORF264	-2	18376..18501	41	catattatcataaagaacaaagtaac	ttg	taa
100997	96ORF265	-2	367..492	41	ctaaacgaaaaagagggtacaatac	atc	tga
100998	96ORF266	-3	32802..32927	41	aggatcatccatttgatacaatact	ttg	taa
100999	96ORF267	-3	10194..10319	41	atcatcgaaaggcgataaactcgtaa	ttg	tga
101000	96ORF268	1	1159..1281	40	ttattcttctcttttgaattgttaa	atg	taa
101001	96ORF269	2	10373..10495	40	gacagagttgaaaagaaaatcatga	atg	taa
101002	96ORF270	2	15734..15856	40	ttattcggcgtaatcgactgatgc	ttg	tag
101003	96ORF271	-1	43451..43573	40	c c tNo shine-dalgarno sequence	att	tga
101004	96ORF272	-1	36959..37081	40	acgctataaaaaaacttttatttag	atg	tag
101005	96ORF273	-1	35798..35920	40	ctgacgcactttgttggtttagtgc	att	taa
101006	96ORF274	-1	8147..8269	40	ctgtctctctatggtttgttagtct	ctg	tga
101007	96ORF275	-2	43066..43188	40	tttaacttactaattttcttttgat	ata	tga
101008	96ORF276	-2	42535..42657	40	aaataatgtaaatgttttcatagt	att	tag
101009	96ORF277	-2	30628..30750	40	ttctagctcccgcttctgcaaaagt	ctg	taa

101010	96ORF278	-2	13291..13413	40	ttcgtatcttccaagcaattcattt	ttg	tga
101011	96ORF279	-2	3172..3294	40	cagattggttagtaacgcctaattt	atc	taa
101012	96ORF280	-3	18804..18926	40	taaataaccaaacacgtgtatcaaca	att	tag
101013	96ORF281	-3	15843..15965	40	atttaaaaaagtgattctataacca	atc	tag
101014	96ORF282	-3	8460..8582	40	ttagtcatcactcaattctttttcc	att	taa
101015	96ORF283	-3	7593..7715	40	gatgttggtctacacagtgctaacac	atg	taa
101016	96ORF284	-3	6453..6575	40	aattcaatttttaattaccattttcta	att	tga
101017	96ORF285	1	15082..15201	39	caatacttagtcacacattcaaaag	att	taa
101018	96ORF286	1	34444..34563	39	acacaaacggttaataagcaaaaagtga	atg	tag
101019	96ORF287	2	27920..28039	39	cctatttttagcaggttggtgcagtaa	ttg	tag
101020	96ORF288	2	28415..28534	39	atcggctttttaactggcgtaattga	atc	tag
101021	96ORF289	2	38147..38266	39	tatcaaatgcttaatttaggcaagt	atc	tga
101022	96ORF290	3	40917..41036	39	gcaaatttaaacactttcacatcat	atg	taa
101023	96ORF291	-2	38815..38934	39	tctctaaaaacagcttacagcgaac	ata	taa
101024	96ORF292	-2	32671..32790	39	ctataggattataaatcgctgacgt	ata	tga
101025	96ORF293	-2	31216..31335	39	ttgatttgatggtttcttatacttga	ttg	taa
101026	96ORF294	-2	21589..21708	39	gtatcttcatcagaatcgccataaa	atc	taa
101027	96ORF295	-2	18976..19095	39	tatcaatatatgctaaccctagcacc	ata	taa
101028	96ORF296	-2	11482..11601	39	gccacctcgactctttttgcaacc	att	taa
101029	96ORF297	-3	12933..13052	39	tcacgaaaataatggtttcttaattt	ata	taa
101030	96ORF298	-3	8262..8381	39	gaactgatcttgcttaaatgattta	att	tag
101031	96ORF299	-3	6993..7112	39	cattagcatttagcgaatgggtttga	ttg	tga
101032	96ORF300	2	23516..23632	38	actacatctgaacaactaaaatttc	atc	tag
101033	96ORF301	2	25943..26059	38	agattagaagaagaaaaaagaagac	gtg	taa
101034	96ORF302	2	36929..37045	38	tattgggggtttgttaacatggggca	atg	tag
101035	96ORF303	3	4476..4592	38	ataaaagctaccttagtagcagta	atg	tga
101036	96ORF304	3	20586..20702	38	tactctaaagatagctaaagcaatac	gtg	tga
101037	96ORF305	3	28356..28472	38	cggttaccaatgtgcttgatacgt	ttg	taa
101038	96ORF306	-1	24359..24475	38	acttaataaaaagccgtatcggtcc	atg	taa
101039	96ORF307	-1	20147..20263	38	ttgtacctatacaggttaactcctt	att	tag
101040	96ORF308	-2	38158..38274	38	ttccgtatccacttttctaagaagac	gtg	tga
101041	96ORF309	-2	35149..35265	38	agcttggtttgtatcgctctttaacga	ata	taa
101042	96ORF310	-2	31423..31539	38	gtaatatgattaggtctcctcttat	ttg	taa
101043	96ORF311	-2	10438..10554	38	cgctttaaactcggttttaggtcact	atc	taa
101044	96ORF312	-2	1390..1506	38	gagaacaacacaaacattaaaca	atc	taa
101045	96ORF313	-3	33051..33167	38	acgtcctggtttcttagatcgtaatac	ata	tag
101046	96ORF314	-3	25194..25310	38	agcaaacccgttaaaagataacattga	atc	taa
101047	96ORF315	-3	6273..6389	38	cattcttgctaacacgtcagattga	ctg	tga
101048	96ORF316	-3	4281..4397	38	ataaactcgattcattaatcattaa	att	tag
101049	96ORF317	1	2260..2373	37	atgactccttttctcatatttcttt	ata	taa
101050	96ORF318	2	21230..21343	37	atttcacacttttttagttgtctct	ttg	taa
101051	96ORF319	3	18018..18131	37	atactgagtcaccaattttaagctcg	atg	tag
101052	96ORF320	3	36972..37085	37	attacagatatcctaaggggtttccg	att	taa
101053	96ORF321	-1	36302..36415	37	ctcttgagtttttttagcttaattta	atc	taa
101054	96ORF322	-1	32606..32719	37	ccataagttatttctccagttcttat	att	taa
101055	96ORF323	-1	11453..11566	37	ctaaacccgttcttttttatcaattc	att	tga
101056	96ORF324	-1	7268..7381	37	tactgggttcgccccagtggaagttct	ata	tga
101057	96ORF325	-2	32347..32460	37	ttactgcattttgtatatggcgataa	atc	tag
101058	96ORF326	-2	24682..24795	37	acgtttattacgctcataaagccat	ata	tag
101059	96ORF327	-2	23905..24018	37	aaatgggtgtggcgcttgaccatat	gtg	taa
101060	96ORF328	-2	21460..21573	37	agagcactaatacgtttttgttctt	ctg	tga
101061	96ORF329	-2	21208..21321	37	gacttaacttcttctgatattcatat	atc	tga
101062	96ORF330	-2	18085..18198	37	ccagtcgacaccagcaaaagtattct	ttg	tag
101063	96ORF331	-2	8170..8283	37	actttgagacgtcgctctgtctctt	atg	tag
101064	96ORF332	-2	5971..6084	37	caatttggttttccggttttctcttag	ttg	tag
101065	96ORF333	-3	37632..37745	37	accttgcttaatacagtcgtaatta	att	tga
101066	96ORF334	-3	29628..29741	37	ctgagtttagtggttgtaaaatgctat	ttg	tag
101067	96ORF335	-3	7164..7277	37	ttagcggatattccgttttcttagtaa	atc	taa
101068	96ORF336	1	22903..23013	36	gtaaaaaaagacaataatgactatta	ctg	tga
101069	96ORF337	1	43258..43368	36	taattgacgtgggttatttttttaggt	ttg	taa
101070	96ORF338	2	12668..12778	36	gaactgggtggaatgggcatggaaca	atc	tag
101071	96ORF339	2	28292..28402	36	ttcaactgcttaattcagttgctta	ctg	taa
101072	96ORF340	2	35396..35506	36	ttcctaatagaacataaagtcaacgggt	att	tga
101073	96ORF341	3	25428..25538	36	actcgagaacaattagaaaaagcaa	ttg	tga
101074	96ORF342	-1	40913..41023	36	tatctgggaatttaacttaataaaa	ata	tga
101075	96ORF343	-1	39173..39283	36	tgccacatttttagtgcaggattga	ttg	tga
101076	96ORF344	-1	37580..37690	36	gggtctaccttttaacgtcgtttcag	ata	taa
101077	96ORF345	-1	31556..31666	36	ggattattcttttctaataaacttcaa	ttg	tga
101078	96ORF346	-1	29972..30082	36	gggtactcttattcttaaaatataat	ttg	tga
101079	96ORF347	-1	28787..28897	36	ctgccaagtcctgtagcaattactt	ttg	tga
101080	96ORF348	-1	21839..21949	36	ttaaaaatccgataaaataacattgc	ctg	tga
101081	96ORF349	-1	3647..3757	36	taaaacttccgaagttaccacagct	ttg	tga

101082	96ORF350	-2	40801..40911	36	accattccaattttgcccataatgat	gcg	tag
101083	96ORF351	-2	38953..39063	36	tatcttttaaaattctcgtaaatagc	atc	taa
101084	96ORF352	-2	31585..31695	36	tagctgtcatcactagatcttttga	atc	taa
101085	96ORF353	-2	24550..24660	36	atagtcggttttaccgcctcgtact	att	tag
101086	96ORF354	-2	20083..20193	36	atcatcattttgatattcttcaaac	ata	tga
101087	96ORF355	-2	991..1101	36	gcattctggcagtcacgacgtaaaac	atc	tag
101088	96ORF356	-3	38148..38258	36	taagaaagcggtgcgcgatcaaatat	att	tga
101089	96ORF357	-3	8790..8900	36	tgaagtctatctagcgctattcttct	ttg	tag
101090	96ORF358	-3	4458..4568	36	ttcataaaagtattcttctgtagtat	atg	tag
101091	96ORF359	1	4666..4773	35	ttatcaaaatatcaacttaatttaa	atc	tag
101092	96ORF360	1	11569..11676	35	ataaatttaccgaacatgaaaatga	att	tga
101093	96ORF361	2	6122..6229	35	ggaaaaaattgatgtgtgtagtga	ttg	taa
101094	96ORF362	-1	40418..40525	35	ttcgtagggtgctacttctttaa	ttg	tag
101095	96ORF363	-1	34358..34465	35	gttttgcttgatttgcatttgttga	atg	tga
101096	96ORF364	-1	20654..20761	35	ctatttccactgattcccatctaa	atg	tga
101097	96ORF365	-1	8423..8530	35	tcttttttagagttacgaggttca	att	tag
101098	96ORF366	-1	2402..2509	35	tgacgtatggcaacatttttagatca	atc	taa
101099	96ORF367	-2	36607..36714	35	aaaataaaaagccagtgccgaagca	ctg	tag
101100	96ORF368	-2	27061..27168	35	caaatacgtcctgcagcgttcaataa	atc	tag
101101	96ORF369	-2	26470..26577	35	atgagttgttaagtttaccctaaat	atc	taa
101102	96ORF370	-2	10327..10434	35	ccgtgccatcttctcgtataagta	ata	taa
101103	96ORF371	-2	8650..8757	35	gggtacgggttgttactgttgatat	atc	taa
101104	96ORF372	-3	14382..14489	35	gttcttttaattgatctactgttaa	att	taa
101105	96ORF373	-3	8151..8258	35	atgtttgttagtctctgtgtagtct	atg	taa
101106	96ORF374	-3	5007..5114	35	aaacgatttaagtggaacattattc	ata	taa
101107	96ORF375	2	30563..30667	34	cgattagaatactttaaaggagac	ttg	tga
101108	96ORF376	-1	19916..20020	34	tctatgtcaggtaatttgcattaa	att	taa
101109	96ORF377	-1	9236..9340	34	ctttctctgttagtaattgttttaa	atc	taa
101110	96ORF378	-1	9026..9130	34	actcttctatctttagttgttttaa	ata	tag
101111	96ORF379	-2	28447..28551	34	cttttctgataataaagtttagtgc	ttg	tga
101112	96ORF380	-3	40329..40433	34	ccattcaccttctttagagattgtga	ttg	tga
101113	96ORF381	-3	39801..39905	34	caaaagatgaaggctttttccatc	ttg	taa
101114	96ORF382	-3	33831..33935	34	atgttcttctgaactcgattaaagt	atc	tga
101115	96ORF383	-3	33687..33791	34	gttattacgtcttaataacttgtgt	gtg	tag
101116	96ORF384	-3	13530..13634	34	tatacgcactagtactgtacactga	ttg	taa
101117	96ORF385	-3	3843..3947	34	tttgattgattgttctagttaagaa	att	taa
101118	96ORF386	1	12256..12357	33	agtcataaagaagtttagcaatgtga	ttg	tag
101119	96ORF387	2	2207..2308	33	tccaagactctttaaactgttaactt	atc	tag
101120	96ORF388	2	2519..2620	33	attgttgaaatttcgattgatctaaa	atg	tga
101121	96ORF389	2	22517..22618	33	agaagtataatgcgtaatgctttag	atg	tag
101122	96ORF390	2	27302..27403	33	ttccaaaattgggctaataagtgtag	ctg	taa
101123	96ORF391	2	32384..32485	33	actaaaaaggttgagaaagctgtag	atg	taa
101124	96ORF392	2	39287..39388	33	aaaaacggtactgtagtatcaatca	atc	tag
101125	96ORF393	3	18153..18254	33	gtagtatatgccgactttgatttga	atg	taa
101126	96ORF394	3	24189..24290	33	tcagacccttaacatttaacaaactag	ttg	tga
101127	96ORF395	-1	15266..15367	33	tcgataatttgcatagctcgtttta	atg	tag
101128	96ORF396	-2	32239..32340	33	ttttagtgaaagcatctagtgttga	ata	tag
101129	96ORF397	-2	16123..16224	33	ttatgtgtgcctatcatattaacaa	ttg	tag
101130	96ORF398	-2	13648..13749	33	tctttaactgaattgtgaatagcat	ttg	tag
101131	96ORF399	-2	10987..11088	33	acttctgtagggtattcttatatcaa	ttg	tga
101132	96ORF400	-2	3382..3483	33	cttactggtaattcttcaaaatttaa	atg	taa
101133	96ORF401	-3	40794..40895	33	ccatagtgtgtgaagtggtttaaact	ttg	taa
101134	96ORF402	-3	39978..40079	33	atattcctaatacactgaacctaa	att	tga
101135	96ORF403	-3	38607..38708	33	atcttcagtgtaaaatcgacagcca	atg	tag
101136	96ORF404	-3	21288..21389	33	cagacaccgtcttaagtcctcttag	ata	taa

Table 11

SEQUENCE INFORMATION FOR PHAGES MATCHING WITH TABLE 1

M32695

Bacteriophage PM2 nuclease cleavage site

gi|166145|gb|M32695|BM2NCS [166145]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M32693

Bacteriophage PM2 Hind III fragment 4

gi|166144|gb|M32693|BM24HIND3 [166144]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M32693

Bacteriophage PM2 Hind III fragment 4

gi|166144|gb|M32693|BM24HIND3 [166144]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M32694

Bacteriophage PM2 Hind III fragment 3

gi|166143|gb|M32694|BM23HIND3 [166143]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

M26134

Bacteriophage PM2 structural protein gene containing purine/pyrimidine rich regions and anti-Z-DNA-IgG binding regions, complete cds

gi|289360|gb|M26134|BM2PROTIV [289360]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

J02452

bacteriophage fi 3'-terminal region rna

gi|215409|gb|J02452|PFITR3 [215409]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

AF020798

Bacteriophage Chp1 genome DNA, complete sequence

gi|217761|dbj|D00624|BCP1 [217761]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 12 protein links, or 1 genome link)

X72793

Clostridium botulinum C phage BONT/C1, ANTP-139, ANTP-33, ANTP-17, ANTP-70 genes and ORF-22

gi|516171|emb|X72793|CBCBONT [516171]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 6 protein links, or 4 nucleotide neighbors)

X51464

Clostridium botulinum D Phage C3 gene for exoenzyme C3

gi|14907|emb|X51464|CBDPE3 [14907]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

D90210

Bacteriophage c-st (from C. botulinum) C1-tox gene for botulinum C1 neurotoxin

gi|217780|dbj|D90210|CSTC1TOX [217780]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

S49407

type D neurotoxin [bacteriophage d-16 phi, host = C. botulinum, type D, CB16, Genomic, 4087 nt]
gi|260238|gb|S49407|S49407 [260238]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

X53370

Bacteriophage phi29 temperature sensitive mutant TS2(98) DNA polymerase gene
gi|15733|emb|X53370|POTS298 [15733]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 7 nucleotide neighbors)

X53371

Bacteriophage phi29 temperature sensitive mutant TS2(24) DNA polymerase gene
gi|15731|emb|X53371|POTS224 [15731]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 7 nucleotide neighbors)

X05973

Bacteriophage phi29 prohead RNA
gi|15680|emb|X05973|POP29PRO [15680]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, or 4 nucleotide neighbors)

V01155

Left end of bacteriophage phi-29 coding for 15 potential proteins Among
these are the terminal protein and the proteins encoded by the genes 1, 2 (sus), 3, and (probably) 4
gi|15659|emb|V01155|POP29B [15659]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 16 protein links, or 16 nucleotide neighbors)

X73097

Bacteriophage phi-29 left origin of replication
gi|312194|emb|X73097|BP29ORIL [312194]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 5 nucleotide neighbors)

M14430

Bacteriophage phi-29 gene-17 gene, complete cds
gi|215321|gb|M14430|P29G17A [215321]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 6 protein links, or 8 nucleotide neighbors)

M14431

Bacteriophage phi-29 gene-16 gene, complete cds
gi|215319|gb|M14431|P29G16A [215319]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 7 nucleotide neighbors)

M20693

Bacteriophage phi-29 DNA, 3' end
gi|215343|gb|M20693|P29REPINB [215343]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 nucleotide neighbors)

M21016

Bacteriophage phi-29 DNA, 5' end
gi|215342|gb|M21016|P29REPINA [215342]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M12456

Bacteriophage phi-29 genes 9, 10 and 11 encoding p9 tail, incomplete, p10 connector, complete, and p11 lower collar, incomplete, respectively
gi|215338|gb|M12456|P29P9 [215338]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 2 nucleotide neighbors)

M14782

Bacillus phage phi-29 head morphogenesis, major head protein, head fiber protein, tail protein, upper collar protein, lower collar protein, pre-neck appendage protein, morphogenesis(13), lysis, morphogenesis(15), encapsidation genes, complete cds
gi|215323|gb|M14782|P29LATE2 [215323]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 11 protein links, or 11 nucleotide neighbors)

M26968

Bacteriophage phi-29 (from *Bacillus subtilis*) proteins p1 delta-1 genes, complete cds, and the sus1(629) mutation
gi|341558|gb|M26968|P29P1D1A [341558]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

J02448

Bacteriophage f1, complete genome
gi|166201|gb|J02448|F1CCG [166201]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 10 protein links, 205 nucleotide neighbors, or 1 genome link)

M24832

Bacteriophage f2 coat protein gene, partial cds
gi|166228|gb|M24832|F2CRNACA [166228]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

J02451

Bacteriophage fd, strain 478, complete genome
gi|215394|gb|J02451|PFDCG [215394]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,5 MEDLINE links, 10 protein links, 204 nucleotide neighbors, or 1 genome link)

M34834

Bacteriophage fr replicase gene, 5' end
gi|166139|gb|M34834|BFRREGRA [166139]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 9 nucleotide neighbors)

M38325

Bacteriophage fr replicase gene, 5' end
gi|166137|gb|M38325|BFRREGR [166137]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 9 nucleotide neighbors)

M35063

Bacteriophage fr coat protein replicase cistron (R region) RNA
gi|166134|gb|M35063|BFRRCRRA [166134]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 3 nucleotide neighbors)

S66567

alpha-atrial natriuretic factor/coat protein=fusion polypeptide [human, bacteriophage fr, expression vector pFAN15, PlasmidSyntheticRecombinant, 510 nt]
gi|435742|gb|S66567|S66567 [435742]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 15 nucleotide neighbors)

X15031

Bacteriophage fr RNA genome

gi|15071|emb|X15031|LEBFRX [15071]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 4 protein links, 9 nucleotide neighbors, or 1 genome link)

U51233

Mus musculus neutralizing anti-RNA-bacteriophage fr immunoglobulin variable region light chain (IgM) mRNA, partial cds

gi|1277150|gb|U51233|MMU51233 [1277150]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 1669 nucleotide neighbors)

U51232

Mus musculus neutralizing anti-RNA-bacteriophage fr immunoglobulin variable region heavy chain (IgM) mRNA, partial cds

gi|1277148|gb|U51232|MMU51232 [1277148]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 1073 nucleotide neighbors)

U02303

Bacteriophage If1, complete genome

gi|3676280|gb|U02303|B2U02303 [3676280]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 10 protein links, or 1 genome link)

V00604

Phage M13 genome

gi|14959|emb|V00604|INM13X [14959]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 10 protein links, or 205 nucleotide neighbors)

A32252

Synthetic bacteriophage M13 protein III probe

gi|1567340|emb|A32252|A32252 [1567340]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

A32251

Synthetic bacteriophage M13 protein III probe

gi|1567339|emb|A32251|A32251 [1567339]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

M12465

Bacteriophage M13 mp10 mutations in lac operon

gi|215210|gb|M12465|M13LACMUT [215210]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 215 nucleotide neighbors)

M24177

Synthetic Bacteriophage M13 (clone M13.SV.B12) SV40 early promoter region DNA

gi|209416|gb|M24177|SYNSVB12 [209416]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M24176

Synthetic Bacteriophage M13 (clone M13.SV.B11) SV40 early promoter region DNA

gi|209415|gb|M24176|SYNSVB11 [209415]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M24175

Synthetic Bacteriophage M13 (clone M13.SV.8) SV40 early promoter region DNA

gi|208806|gb|M24175|SYNM13SV8 [208806]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 242 nucleotide neighbors)

M19979

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33

gi|207813|gb|M19979|SYN33M13M [207813]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 617 nucleotide neighbors)

M19565

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33

gi|207808|gb|M19565|SYN33M13H [207808]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 567 nucleotide neighbors)

M19564

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33

gi|207807|gb|M19564|SYN33M13G [207807]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 12 nucleotide neighbors)

M19563

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33

gi|207806|gb|M19563|SYN33M13F [207806]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 262 nucleotide neighbors)

M19561

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33

gi|207804|gb|M19561|SYN33M13D [207804]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 27 nucleotide neighbors)

M19560

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33

gi|207803|gb|M19560|SYN33M13C [207803]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

M19559

Synthetic hybrids; recombinant DNA from bacteriophage M13 and plasmid pHV33

gi|207802|gb|M19559|SYN33M13B [207802]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 227 nucleotide neighbors)

M10568

Bacteriophage M13 replicative form II, replication origin, specific nick location

gi|215220|gb|M10568|M13ORIB [215220]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 650 nucleotide neighbors)

M10910

Bacteriophage M13 gene II regulatory region and M13sj1 mutant

gi|215209|gb|M10910|M13IIREG [215209]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 72 nucleotide neighbors)

M38295

Bacteriophage M13 HaeIII restriction fragment DNA

gi|215208|gb|M38295|M13HAEIII [215208]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 67 nucleotide neighbors)

E02067

DNA encoding a part of Bacteriophage M13 tg 127
 gi|2170311|dbj|E02067|E02067 [2170311]
 (View GenBank report, FASTA report, ASN.1 report, or Graphical view)

J02467

Bacteriophage MS2, complete genome
 gi|215232|gb|J02467|MS2CG [215232]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 8 MEDLINE links, 4 protein links, 20 nucleotide neighbors, or 1 genome link)

AJ004950

Bacteriophage P1 ban gene
 gi|3688226|emb|AJ011592|BP1011592 [3688226]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 protein link)

U88974

Bacteriophage P1 structural lytic transglycosylase (orf47), pep44b (orf44b),
 pep44a (orf44a), and pep43 (orf43) genes, complete cds; and pep42 (orf42) gene, partial cds
 gi|2661099|gb|AF035607|AF035607 [2661099]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 5 protein links, or 1 nucleotide neighbor)

AJ000741

Bacteriophage P1 darA operon
 gi|2462938|emb|AJ000741|BPAJ7641 [2462938]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 10 protein links, or 31 nucleotide neighbors)

X01828

Bacteriophage P1 recombinase gene cin
 gi|15133|emb|X01828|MYP1CIN [15133]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

X98146

Bacteriophage P1 DNA sequence around the Op88 operator
 gi|1359513|emb|X98146|BP1OP88OP [1359513]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 nucleotide neighbor)

S61175

imm1 operon: icd=cell division repressor, ant1=antirepressor (promoters
 P51a, P51b) [bacteriophage P1, Genomic, 728 nt]
 gi|385908|gb|S61175|S61175 [385908]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 3 nucleotide neighbors)

X87824

Bacteriophage P1 gene 26
 gi|861164|emb|X87824|XXBP1G26 [861164]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 protein link)

X15638

Phage P1 DNA for lytic replicon containing promoter P53 and two open reading frames
 gi|15735|emb|X15638|PP1LREP [15735]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 24 nucleotide neighbors)

X17512

Bacteriophage P1 DNA for immunity region immI

gi|15479|emb|X17512|P1IMMUNITY [15479]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, or 4 nucleotide neighbors)

X16005

Bacteriophage P1 cI gene for P1cI repressor protein

gi|15477|emb|X16005|P1C1 [15477]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

X03453

Bacteriophage P1 cre gene for recombinase protein

gi|15135|emb|X03453|MYP1CRE [15135]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 12 nucleotide neighbors)

X06561

Bacteriophage P1 cI gene 5'-region

gi|15128|emb|X06561|MYP1C1 [15128]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 6 nucleotide neighbors)

V01534

Bacteriophage P1 genome fragment (IS2 insertion spot). This regions contains

four unidentified reading frames and is known as insertion hot spot for IS2 insertion sequences

gi|15118|emb|V01534|MYOVPI [15118]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 3 nucleotide neighbors)

X56951

Bacteriophage P1 gene10

gi|406728|emb|X56951|BPP1GP10 [406728]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 3 protein links, or 1 nucleotide neighbor)

K02380

Bacteriophage P1 replication region including repA, parA, and parB genes and

incA, incB, and incC incompatibility determinants

gi|215652|gb|K02380|PP1REP [215652]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,5 MEDLINE links, 4 protein links, or 8 nucleotide neighbors)

X87674

Bacteriophage P1 lydA & lydB genes

gi|974763|emb|X87674|BACP1LYD [974763]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

X87673

Bacteriophage P1 gene 17

gi|974761|emb|X87673|BACP117 [974761]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

M16618

Bacteriophage P1 cI repressor binding sites

gi|215600|gb|M16618|PP1C1 [215600]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 3 nucleotide neighbors)

SEG_PP1CIN

Bacteriophage P1 cin gene encoding recombinase, cixL recombination site, and 5' end of C invertible element

gi|215607|gb|SEG_PP1CIN [215607]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

K03173

Bacteriophage P1 C invertible element, right end, and cixR recombination site

gi|215606|gb|K03173|PP1CIN2 [215606]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

215605

Bacteriophage P1 cin gene encoding recombinase, cixL recombination site, and 5' end of C invertible element

gi|215605|cl|X01828 [215605]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

M25470

Bacteriophage P1 tail fiber protein gene, complete cds

gi|341349|gb|M25470|PP1TFPR [341349]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 3 nucleotide neighbors)

M34382

Bacteriophage P1 sim region proteins, complete cds

gi|215661|gb|M34382|PP1SIM [215661]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

M81956

Bacteriophage P1 R protein (R) gene, complete cds

gi|215658|gb|M81956|PP1RP [215658]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 4 nucleotide neighbors)

M37080

Bacteriophage P1 mini-P1 plasmid origin of replication

gi|215657|gb|M37080|PP1REPOR [215657]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 46 nucleotide neighbors)

M27041

Bacteriophage P1 ref gene, complete cds

gi|215650|gb|M27041|PP1REF [215650]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

L01408

Bacteriophage P1 partition protein (parB) gene, 3' end

gi|215642|gb|L01408|PP1PARB [215642]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 41 nucleotide neighbors)

SEG_PP1PAR

Bacteriophage miniplasmid P1 parA gene, 5' end

gi|215639|gb|SEG_PP1PAR [215639]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 48 nucleotide neighbors)

M36425

Bacteriophage miniplasmid P1 parB gene, 3' end

gi|215638|gb|M36425|PP1PAR2 [215638]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

M36424

Bacteriophage miniplasmid P1 parA gene, 5' end
gi|215637|gb|M36424|PP1PAR1 [215637]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

M11129

Bacteriophage P1 miniplasmid origin of replication region
gi|215632|gb|M11129|PP1ORIM [215632]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 43 nucleotide neighbors)

M25414

Bacteriophage P1 c1 repressor binding site, operator 88 (Op88)
gi|215631|gb|M25414|PP1OP88A [215631]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 3 nucleotide neighbors)

M25413

Bacteriophage P1 c1 repressor binding site, operator 68 (Op68)
gi|215630|gb|M25413|PP1OP68A [215630]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

M25412

Bacteriophage P1 c1 repressor binding site, operator 21 (Op21)
gi|215629|gb|M25412|PP1OP21A [215629]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M10510

Bacteriophage P1 recombination site loxR
gi|215628|gb|M10510|PP1LOXR [215628]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M10287

Bacteriophage P1 loxP X loxP recombination site
gi|215627|gb|M10287|PP1LOXPX [215627]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 13 nucleotide neighbors)

M10494

Bacteriophage P1 recombination site loxP
gi|215626|gb|M10494|PP1LOXP [215626]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 134 nucleotide neighbors)

M10511

Bacteriophage P1 recombination site loxL
gi|215625|gb|M10511|PP1LOXL [215625]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

M10512

Bacteriophage P1 recombination site loxB
gi|215624|gb|M10512|PP1LOXB [215624]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

M10145

Bacteriophage P1 genome fragment with recombination site loxP
gi|215623|gb|M10145|PP1CREX [215623]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 21 nucleotide neighbors)

M13327

Bacteriophage P1 *Cin* recombinase activated cross over site, junction IV, clone pSHI326
gi|215622|gb|M13327|PP1CN26IV [215622]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

M13325

Bacteriophage P1 *Cin* recombinase activated cross over site, junction II, clone pSHI326
gi|215621|gb|M13325|PP1CN26II [215621]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1401 nucleotide neighbors)

M13323

Bacteriophage P1 *Cin* recombinase activated cross over site, junction IV, clone pSHI325
gi|215620|gb|M13323|PP1CN25IV [215620]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

M13321

Bacteriophage P1 *Cin* recombinase activated cross over site, junction II, clone pSHI325
gi|215619|gb|M13321|PP1CN25II [215619]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1058 nucleotide neighbors)

M13324

Bacteriophage P1 *Cin* recombinase activated cross over site, junction I, clone pSHI326
gi|215618|gb|M13324|PP1CIR26I [215618]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

M13319

Bacteriophage P1 *Cin* recombinase activated cross over site, right junction, clone pSHI327
gi|215617|gb|M13319|PP1CIN27R [215617]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

M13320

Bacteriophage P1 *Cin* recombinase activated cross over site, junction I, clone pSHI325
gi|215616|gb|M13320|PP1CIN25I [215616]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

M13318

Bacteriophage P1 *Cin* recombinase activated cross over site, left junction, clone pSHI324
gi|215615|gb|M13318|PP1CIN24L [215615]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1370 nucleotide neighbors)

M13317

Bacteriophage P1 *Cin* recombinase activated cross over site, right junction, clone pSHI323
gi|215614|gb|M13317|PP1CIN23M [215614]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1055 nucleotide neighbors)

M13316

Bacteriophage P1 *Cin* recombinase activated cross over site, left junction, clone pSHI323
gi|215613|gb|M13316|PP1CIN23L [215613]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

M13315

Bacteriophage P1 *Cin* recombinase activated cross over site, right junction, clone pSHI322
gi|215612|gb|M13315|PP1CIN22R [215612]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 7 nucleotide neighbors)

215

M13314

Bacteriophage P1 *Cin* recombinase activated cross over site, left junction, clone pSHI322

gi|215611|gb|M13314|PP1CIN22L [215611]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1401 nucleotide neighbors)

M13313

Bacteriophage P1 *Cin* recombinase activated cross over site, right junction, clone pSHI321

gi|215610|gb|M13313|PP1CIN21R [215610]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 7 nucleotide neighbors)

M13312

Bacteriophage P1 *Cin* recombinase activated cross over site, left junction, clone pSHI321

gi|215609|gb|M13312|PP1CIN21L [215609]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1058 nucleotide neighbors)

M16568

Bacteriophage P1 *c4* repressor gene, complete cds

gi|215603|gb|M16568|PP1C4 [215603]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

M13326

Bacteriophage P1 *Cin* recombinase activated cross over site, junction III, clone pSHI326

gi|215602|gb|M13326|PP1C26III [215602]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1192 nucleotide neighbors)

M13322

Bacteriophage P1 *Cin* recombinase activated cross over site, junction III, clone pSHI325

gi|215601|gb|M13322|PP1C25III [215601]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1231 nucleotide neighbors)

J05651

Bacteriophage P1 modulator protein (*bof*) gene, complete cds

gi|215598|gb|J05651|PP1BOFY1 [215598]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

M33224

Bacteriophage P1 regulatory protein (*bof*) gene, complete cds

gi|215596|gb|M33224|PP1BOFFO [215596]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

M10288

E.coli/bacteriophage P1 *loxR* recombination site

gi|146647|gb|M10288|ECOLOXR [146647]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 3 nucleotide neighbors)

M10289

E.coli/bacteriophage P1 *loxL* recombination site

gi|146646|gb|M10289|ECOLOXL [146646]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

M10290

E.coli *loxB* site, which can recombine with bacteriophage P1 *loxP* site

gi|146645|gb|M10290|ECOLOXB [146645]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

M10287

Bacteriophage P1 loxP X loxP recombination site

gi|215627|gb|M10287|PP1LOXPX [215627]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 13 nucleotide neighbors)

M74046

Bacteriophage P1 pacA and pacB genes, complete cds

gi|215634|gb|M74046|PP1PACAB [215634]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

M95666

Bacteriophage P1 gene 10, doc and phd genes, complete cds

gi|463276|gb|M95666|PP1PHDDOC [463276]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 4 protein links, or 1 nucleotide neighbor)

M25604

Bacteriophage Q-beta mutated autonomously replicating sequence MDV1 RNA fragment

gi|556359|gb|M25604|PQBARSMUT [556359]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 8 nucleotide neighbors)

V00643

first half of the phage Q-beta gene for coat protein

gi|15088|emb|V00643|LEQBET [15088]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

M25167

Bacteriophage Q-beta RNA fragment recovered from replicase binding complex

gi|556362|gb|M25167|PQBREPLICB [556362]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 nucleotide neighbors)

M24876

Bacteriophage Q-beta replicase RNA, 5' end

gi|556360|gb|M24876|PQBREPLICA [556360]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

M25444

Synthetic bacteriophage Q-beta DNA

gi|209118|gb|M25444|SYNPQBTERM [209118]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 8 nucleotide neighbors)

M25463

Bacteriophage Q-beta self-replicating microvariant (+) RNA

gi|532489|gb|M25463|PQBMVSRRNA [532489]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

M25014

Bacteriophage Q-beta RNA replicase gene, 5' end, and maturation protein gene, 3' end

gi|294316|gb|M25014|PQBREPLC [294316]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

M25065

Bacteriophage Q-beta RNA sequence with putative stem loop

gi|294315|gb|M25065|PQBLLOOP [294315]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 3 nucleotide neighbors)

M10265

Bacteriophage Q-beta RNA molecule with the ability to replicate extracellularly

gi|215726|gb|M10265|PQBRNA [215726]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 8 nucleotide neighbors)

M24815

Bacteriophage Q-beta specified replicase subunit RNA,

gi|215725|gb|M24815|PQBREPL [215725]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 nucleotide neighbors)

M25461

Bacteriophage Q-beta plus-strand RNA, 5' terminus

gi|215724|gb|M25461|PQBPSSE [215724]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

M25462

Bacteriophage Q-beta plus-strand RNA, 3' terminus

gi|215723|gb|M25462|PQBPS3E [215723]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 8 nucleotide neighbors)

M24871

Bacteriophage Q-beta nanovariant WSIII RNA

gi|215722|gb|M24871|PQBNVWSIC [215722]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 nucleotide neighbors)

M24870

Bacteriophage Q-beta nanovariant WSII RNA

gi|215721|gb|M24870|PQBNVWSIB [215721]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 nucleotide neighbors)

M24869

Bacteriophage Q-beta nanovariant WSI RNA

gi|215720|gb|M24869|PQBNVWSIA [215720]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 nucleotide neighbors)

M10495

Coliphage Q-beta MDV-1(+) RNA

gi|215719|gb|M10495|PQBMDV1A [215719]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 10 nucleotide neighbors)

J02484

bacteriophage qbeta coat protein cistron first half

gi|215717|gb|J02484|PQBPCP5 [215717]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

M57754

Bacteriophage Q-beta minus strand RNA, 5' terminus

gi|215716|gb|M57754|PQBBSSE [215716]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 8 nucleotide neighbors)

M24297

Bacteriophage Q-beta 5'-terminal region of the minus strand

gi|215715|gb|M24297|PQB5END [215715]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 8 nucleotide neighbors)

M10695

218

Bacteriophage Q-beta, MDV-1 RNA

gi|215714|gb|M10695|PQB1IR [215714]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, or 12 nucleotide neighbors)

M24827

Bacteriophage R17 A protein gene, 5' end

gi|216078|gb|M24827|R17RNACIS [216078]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 5 nucleotide neighbors)

M24829

Bacteriophage R17 coat protein gene, 5' end

gi|216075|gb|M24829|R17CP5 [216075]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 5 nucleotide neighbors)

J02488

bacteriophage r17 rna synthetase initiation site

gi|216080|gb|J02488|R17RNASYN [216080]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 2 protein links, or 6 nucleotide neighbors)

J02487

bacteriophage r17 coat protein initiation site

gi|216073|gb|J02487|R17COATP [216073]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

J02486

bacteriophage r17 a protein initiation site

gi|216071|gb|J02486|R17APROT [216071]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

M24826

Bacteriophage R17 coat protein RNA fragment

gi|216077|gb|M24826|R17CPRAA [216077]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 7 nucleotide neighbors)

M24296

Bacteriophage R17 3'-terminal fragment A RNA

gi|216070|gb|M24296|R173TFA [216070]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 9 nucleotide neighbors)

1TFN

structure refinement for a 24-nucleotide rna hairpin, nmrx, minimized average

structure ribonucleic acid, hairpin, bacteriophage r17 mol_id: 1; molecule: r17c; chain: null; engineered: yes

gi|1942336|pdb|1TFN| [1942336]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 structure link)

1RPEA

rna (5'-d(gp gpgpapcpupgpgpapcpupcpapcp g pcpagpupcpupapu-3') (24-mer rna

hairpin coat protein binding site for bacteriophage r17) (nmrx, minimized average structure)

gi|1421020|pdb|1RHT| [1421020]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 structure link)

M14428

Bacteriophage S13 circular DNA, complete genome

gi|216089|gb|M14428|S13CG [216089]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 12 protein links, 26 nucleotide neighbors, or 1 genome link)

J05393

Bacteriophage T1 DNA N-6-adenine-methyltransferase (M.T1) gene, complete cds

gi|166163|gb|J05393|BT1NAMTA [166163]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 protein links)

L46845

Bacteriophage T2 frd3, frd2 genes, complete cds

gi|951387|gb|L46845|PT2FRD32G [951387]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 17 nucleotide neighbors)

L43611

Bacteriophage T2 fibrin (wac) gene, complete cds

gi|903869|gb|L43611|PT2WAC [903869]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 4 nucleotide neighbors)

M24812

Bacteriophage T2 secondary structure RNA sequence

gi|215796|gb|M24812|PT2RNA [215796]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 4 nucleotide neighbors)

M22342

Bacteriophage T2 DNA-(adenine-N6)methyltransferase (dam) gene, complete cds

gi|215792|gb|M22342|PT2DAM [215792]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

S57515

orf 61.2 {intergenic region between 41 and 61} [bacteriophage T2, Genomic, 323 nt]

gi|298524|gb|S57515|S57515 [298524]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

X05312

Bacteriophage T2 gene 38 for receptor recognizing protein

gi|15197|emb|X05312|MYT2G38 [15197]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

X04442

Bacteriophage T2 gene 37 for receptor recognizing protein

gi|15195|emb|X04442|MYT2G37 [15195]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

X12460

Bacteriophage T2 gene 32 mRNA for single-stranded DNA binding protein

gi|15192|emb|X12460|MYT2G32 [15192]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 14 nucleotide neighbors)

X57797

Bacteriophage T2 gene for gp12

gi|14875|emb|X56555|BT2GP12 [14875]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 2 nucleotide neighbors)

X01755

Bacteriophage T2 tail fiber gene 36

gi|15189|emb|X01755|MYT2F36 [15189]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

M14784

Bacteriophage T3 strain amNG220B right end, tail fiber protein, lysis protein and DNA packaging proteins, complete cds

gi|215810|gb|M14784|PT3RE [215810]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 9 protein links, or 10 nucleotide neighbors)

SEG_PT3RNAPOL

Bacteriophage T3 RNA polymerase III gene, 5' end

gi|710559|gb|SEG_PT3RNAPOL [710559]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

M22610

Bacteriophage T3 RNA polymerase III gene, 3' end

gi|340722|gb|M22610|PT3RNAPOL2 [340722]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M22609

Bacteriophage T3 RNA polymerase III gene, 5' end

gi|340721|gb|M22609|PT3RNAPOL1 [340721]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

X05031

Bacteriophage T3 gene region 1-2.5 with primary origin of replication

gi|15719|emb|X05031|POT3ORI [15719]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 11 protein links, or 5 nucleotide neighbors)

X03964

Bacteriophage T3 early control region pos. 308-810 from genome left end

gi|15718|emb|X03964|POT3EP [15718]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, or 20 nucleotide neighbors)

X17255

Bacteriophage T3 gene 1 to gene 11

gi|15682|emb|X17255|POT3111G [15682]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,4 MEDLINE links, 36 protein links, 17 nucleotide neighbors, or 1 genome link)

X15840

Phage T3 gene 10

gi|15625|emb|X15840|PODT3G10 [15625]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 3 nucleotide neighbors)

X02981

Bacteriophage T3 gene 1 for RNA polymerase

gi|15561|emb|X02981|PODOT3P [15561]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

J02503

bacteriophage t3 5' end, terminally redundant sequence (trs)

gi|215816|gb|J02503|PT3TRS1 [215816]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

SEG_PT3TRS

bacteriophage λ 5' end, terminally redundant sequence (trs)

gi|215818|gb|SEG_PT3TRS [215818]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

J02504

bacteriophage λ 3' end, terminally redundant sequence (trs)

gi|215817|gb|J02504|PT3TRS2 [215817]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

HYPERLINK <http://www.rs.noda.sut.ac.jp/~kunisawa> <http://www.rs.noda.sut.ac.jp/~kunisawa>

Bacteriophage T4 genomic database compiled by Arisaka et al.

X95646

Bacteriophage T5 DNA for region 60.5%-71% of the T5 genome

gi|2791557|emb|AJ001191|BTJ001191 [2791557]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 7 MEDLINE links, 12 protein links, or 6 nucleotide neighbors)

X56847

Bacteriophage T5 genomic region encoding early genes D10-D15

gi|15407|emb|X12930|MYT5D10 [15407]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 5 protein links, or 4 nucleotide neighbors)

AF039886

Bacteriophage T5 subclone T5.5.3r5.18r, single pass sequence, genomic survey sequence

gi|2811154|gb|AF039886|AF039886 [2811154]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039885

Bacteriophage T5 subclone T5.40f,41f, single pass sequence, genomic survey sequence

gi|2811153|gb|AF039885|AF039885 [2811153]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039884

Bacteriophage T5 subclone T5.26.fr, single pass sequence, genomic survey sequence

gi|2811152|gb|AF039884|AF039884 [2811152]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039883

Bacteriophage T5 subclone 10-T5.5.7F, single pass sequence, genomic survey sequence

gi|2811151|gb|AF039883|AF039883 [2811151]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039882

Bacteriophage T5 subclone 41-T5.5.4BF, single pass sequence, genomic survey sequence

gi|2811150|gb|AF039882|AF039882 [2811150]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF039881

Bacteriophage T5 subclone 39-T5.5.4aF, single pass sequence, genomic survey sequence

gi|2811149|gb|AF039881|AF039881 [2811149]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 nucleotide neighbor)

AF039880

Bacteriophage T5 subclone 19-T5.7.2r, single pass sequence, genomic survey sequence
gi|2811148|gb|AF039880|AF039880 [2811148]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF039879

Bacteriophage T5 subclone 18-T5.7.2F, single pass sequence, genomic survey sequence
gi|2811147|gb|AF039879|AF039879 [2811147]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF039878

Bacteriophage T5 subclone 11-T5.5.7R, single pass sequence, genomic survey sequence
gi|2811146|gb|AF039878|AF039878 [2811146]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 2
nucleotide neighbors)

AF039877

Bacteriophage T5 subclone T5.4FR, single pass sequence, genomic survey sequence
gi|2811145|gb|AF039877|AF039877 [2811145]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF039876

Bacteriophage T5 subclone 22-T5.16R, single pass sequence, genomic survey sequence
gi|2811144|gb|AF039876|AF039876 [2811144]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF039875

Bacteriophage T5 subclone 21-T5.16R, single pass sequence, genomic survey sequence
gi|2811143|gb|AF039875|AF039875 [2811143]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF039874

Bacteriophage T5 subclone 21-T5.16F, single pass sequence, genomic survey sequence
gi|2811142|gb|AF039874|AF039874 [2811142]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF039873

Bacteriophage T5 subclone 09-T5.6F, single pass sequence, genomic survey sequence
gi|2811141|gb|AF039873|AF039873 [2811141]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF039872

Bacteriophage T5 subclone 09-T5.6R, single pass sequence, genomic survey sequence
gi|2811140|gb|AF039872|AF039872 [2811140]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 2 nucleotide neighbors)

AF039871

Bacteriophage T5 subclone 04-T5.26.R, single pass sequence, genomic survey sequence
gi|2811139|gb|AF039871|AF039871 [2811139]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF039870

Bacteriophage T5 subclone 13-T5.42F, single pass sequence, genomic survey sequence
gi|2811138|gb|AF039870|AF039870 [2811138]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

X69460

Bacteriophage T5 ltf gene for L-shaped tail fibers

gi|15415|emb|X69460|MYT5LTF [15415]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 1 protein link, or 4 nucleotide neighbors)

X03402

Bacteriophage T5 D15 gene for 5' exonuclease

gi|15413|emb|X03402|MYT5EXOG [15413]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

Z11972

Bacteriophage T5 tRNA-Tyr, tRNA-Glu, tRNA-Trp, tRNA-Phe, tRNA-Cys and tRNA-Asn genes, and ORFs 91aa, 90aa, 42aa and 172aa

gi|15795|emb|Z11972|T56TRNAG [15795]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 3 nucleotide neighbors)

X03898

Bacteriophage T5 genes for tRNA-His, -Ser and -Leu

gi|15786|emb|X03898|STT5RN1 [15786]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 2 MEDLINE links)

X04177

Bacteriophage T5 gene for transfer RNA-Gln

gi|15421|emb|X04177|MYT5TRNQ [15421]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

X03899

Bacteriophage T5 genes for tRNA-Val, -Lys, -fMet, -Pro and -Ile3

gi|15787|emb|X03899|STT5RN2 [15787]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

X03798

Bacteriophage T5 gene for tRNA-Asp (GUC)

gi|15472|emb|X03798|NCT5TRDG [15472]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

Y00364

Bacteriophage T5 tRNA gene cluster (27.8%-22.4%)

gi|15420|emb|Y00364|MYT5TRN [15420]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 13 nucleotide neighbors)

X03140

Bacteriophage T5 DNA with rho-dependent transcription terminator (Hind III-P fragment)

gi|15417|emb|X03140|MYT5RHO [15417]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

Z35070

Bacteriophage T6 DNA

gi|535228|emb|Z35074|MYEREGBT6 [535228]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

AF060870

Coliphage T6 small subunit distal tail fiber (gene 36) gene, partial cds; and large subunit distal tail fiber (gene 37) and tail fiber adhesin (gene 38) genes, complete cds

gi|3676458|gb|AF052605|AF052605 [3676458]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 protein links, or 2 nucleotide neighbors)

Z35072

Bacteriophage T6 DNA encoding ORF19.1 gene and g19 gene

gi|535232|emb|Z35072|MYTAILT6 [535232]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 protein links)

X12488

Bacteriophage T6 gene 32 mRNA for single-stranded DNA binding protein

gi|15843|emb|X12488|MYT6G32 [15843]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 14 nucleotide neighbors)

Z78095

Bacteriophage T6 DNA (1506 bp)

gi|1488562|emb|Z78095|BPHZ78095 [1488562]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 4 nucleotide neighbors)

Z35079

Bacteriophage T6 DNA for Ip5, Ip6

gi|535215|emb|Z35079|MY57BT6 [535215]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

X68725

E.coli bacteriophage T6 gene for beta-glucosyl-HMC-alpha-glucosyl-transferase

gi|296439|emb|X68725|ECT6 [296439]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

X69894

Bacteriophage T6 alt gene for ADP-Ribosyltransferase

gi|15422|emb|X69894|MYT6ADP [15422]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

L46846

Bacteriophage T6 frd3, frd2 genes, complete cds

gi|951390|gb|L46846|PT6FRD32G [951390]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 2 protein links)

M27738

Bacteriophage T6 translational repressor protein (regA), complete cds

gi|215993|gb|M27738|PT6REGA [215993]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 5 nucleotide neighbors)

M38465

Bacteriophage T6 DNA ligase gene, complete cds

gi|215991|gb|M38465|PT6LIG55 [215991]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

V01146

Genome of bacteriophage T7

gi|431187|emb|V01146|T7CG [431187]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 13 MEDLINE links, 60 protein links, 105 nucleotide neighbors, or 1 genome link)

X60322

Bacteriophage alpha3 genes A, B, K, C, D, E, J, F, G, H

gi|14775|emb|X60322|BACALPHA [14775]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 10 protein links, 22 nucleotide neighbors, or 1 genome link)

X13332

Bacteriophage alpha3 DNA for origin of replication

gi|15093|emb|X13332|MLA3ORPL [15093]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

X12611

Bacteriophage alpha3 gene for protein A part, finger domain

gi|15092|emb|X12611|MLA3AFIN [15092]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 6 nucleotide neighbors)

X15721

Bacteriophage alpha3 deletion mutation DNA for the origin region (-ori) of replication

gi|14774|emb|X15721|BA3DMOR9 [14774]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 11 nucleotide neighbors)

X15720

Bacteriophage alpha3 deletion mutant DNA for the origin region (-ori) of replication

gi|14773|emb|X15720|BA3DMOR8 [14773]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 nucleotide neighbor)

X15719

Bacteriophage alpha3 insertion mutant DNA for the origin region (-ori) of replication

gi|14772|emb|X15719|BA3DMOR7 [14772]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 10 nucleotide neighbors)

X15718

Bacteriophage alpha3 deletion mutation DNA for origin region (-ori) of replication

gi|14771|emb|X15718|BA3DMOR6 [14771]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 11 nucleotide neighbors)

X15717

Bacteriophage alpha3 deletion mutant DNA for origin region (-ori) of replication

gi|14770|emb|X15717|BA3DMOR5 [14770]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 9 nucleotide neighbors)

X15716

Bacteriophage alpha3 deletion mutant DNA for origin region (-ori) of replication

gi|14769|emb|X15716|BA3DMOR4 [14769]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 10 nucleotide neighbors)

X15715

Bacteriophage alpha3 deletion mutant DNA for origin region (-ori) of replication

gi|14768|emb|X15715|BA3DMOR3 [14768]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 11 nucleotide neighbors)

X15714

Bacteriophage alpha3 deletion mutant DNA for origin region (-ori) of replication

gi|14767|emb|X15714|BA3DMOR2 [14767]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 11 nucleotide neighbors)

X15713

Bacteriophage alpha3 deletion mutant DNA for the origin region (-ori) of replication

gi|14766|emb|X15713|BA3DMOR1 [14766]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 11 nucleotide neighbors)

X62059

Bacteriophage alpha3 origin of cDNA synthesis (oriGA)

gi|14763|emb|X62059|AL3ORIGA [14763]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 13 nucleotide neighbors)

X62058

Bacteriophage alpha3 origin of cDNA synthesis (oriAA)

gi|14762|emb|X62058|AL3ORIAA [14762]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 13 nucleotide neighbors)

J02444

Bacteriophage alpha3 origin of DNA replication

gi|166103|gb|J02444|AL3ORI [166103]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 12 nucleotide neighbors)

M25640

Bacteriophage alpha-3 H protein gene, complete cds

gi|166101|gb|M25640|AL3HP [166101]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 13 nucleotide neighbors)

M10631

Bacteriophage alpha-3 cleavage site for phage phi-X174 gene A protein

gi|166099|gb|M10631|AL3CSA [166099]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

X00774

Bacteriophage alpha-3 gene J sequence

gi|15431|emb|X00774|NCBA3J [15431]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 2 nucleotide neighbors)

M25640

Bacteriophage alpha-3 H protein gene, complete cds

gi|166101|gb|M25640|AL3HP [166101]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 13 nucleotide neighbors)

M10631

Bacteriophage alpha-3 cleavage site for phage phi-X174 gene A protein

gi|166099|gb|M10631|AL3CSA [166099]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

J02459

Bacteriophage lambda, complete genome

gi|215104|gb|J02459|LAMCG [215104]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,87 MEDLINE links, 67 protein links, 190 nucleotide neighbors, or 1 genome link)

J02482

Bacteriophage phi-X174, complete genome

gi|216019|gb|J02482|PX1CG [216019]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,23 MEDLINE links, 11 protein links, 26 nucleotide neighbors, or 1 genome link)

J02454

Bacteriophage G4, complete genome

gi|215415|gb|J02454|PG4CG [215415]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,6 MEDLINE links, 11 protein links, 20 nucleotide neighbors, or 1 genome link)

X60323

Bacteriophage phiK complete genome

gi|1478118|emb|X60323|BPHKCG [1478118]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,10 protein links, 18 nucleotide neighbors, or 1 genome link)

L42820

Bacteriophage BF23 tail protein (hrs) gene, complete cds

gi|1048680|gb|L42820|BBFHRS [1048680]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

X54455

Bacteriophage BF23 gene 17 and gene 18

gi|14797|emb|X54455|BF231718G [14797]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 2 nucleotide neighbors)

M37097

Bacteriophage BF23 DNA, right end of terminal repetition

gi|166115|gb|M37097|BBFRIGH [166115]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

M37096

Bacteriophage BF23 DNA, left end of terminal repetition

gi|166114|gb|M37096|BBFLEFT [166114]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 nucleotide neighbor)

M37095

Bacteriophage BF23 A2-A3 gene, complete cds, and A1 gene, 5' end

gi|166110|gb|M37095|BBFA2A3 [166110]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 3 protein links, or 1 nucleotide neighbor)

AF056281

Bacteriophage BF23 clone b23.mac5/6.1, genomic survey sequence

gi|3090930|gb|AF056281|AF056281 [3090930]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056280

Bacteriophage BF23 clone bf23.mac3, genomic survey sequence
gi|3090929|gb|AF056280|AF056280 [3090929]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056279

Bacteriophage BF23 clone bf23.mac18/21.34, genomic survey sequence
gi|3090928|gb|AF056279|AF056279 [3090928]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056278

Bacteriophage BF23 clone bf23.mac16/19.33, genomic survey sequence
gi|3090927|gb|AF056278|AF056278 [3090927]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056277

Bacteriophage BF23 clone bf23.mac16/19-33, genomic survey sequence
gi|3090926|gb|AF056277|AF056277 [3090926]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056276

Bacteriophage BF23 clone bf23.mac12/9-9, genomic survey sequence
gi|3090925|gb|AF056276|AF056276 [3090925]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056275

Bacteriophage BF23 clone bf23.mac11/14-24, genomic survey sequence
gi|3090924|gb|AF056275|AF056275 [3090924]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056274

Bacteriophage BF23 clone bf23.57r64r, genomic survey sequence
gi|3090923|gb|AF056274|AF056274 [3090923]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 3 nucleotide neighbors)

AF056273

Bacteriophage BF23 clone bf23.54fr, genomic survey sequence
gi|3090922|gb|AF056273|AF056273 [3090922]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056272

Bacteriophage BF23 clone bf23.47fr.mac10/7, genomic survey sequence
gi|3090921|gb|AF056272|AF056272 [3090921]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056271

Bacteriophage BF23 clone bf23.23.66r, genomic survey sequence
gi|3090920|gb|AF056271|AF056271 [3090920]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056270

Bacteriophage BF23 clone bf23.23.64f, genomic survey sequence
gi|3090919|gb|AF056270|AF056270 [3090919]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056269

Bacteriophage BF23 clone bf23.23.60r, genomic survey sequence
gi|3090918|gb|AF056269|AF056269 [3090918]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056268

Bacteriophage BF23 clone bf23.23.60f, genomic survey sequence
gi|3090917|gb|AF056268|AF056268 [3090917]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 nucleotide neighbor)

AF056267

Bacteriophage BF23 clone bf23.23.59r, genomic survey sequence
gi|3090916|gb|AF056267|AF056267 [3090916]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056266

Bacteriophage BF23 clone bf23.23.59f, genomic survey sequence
gi|3090915|gb|AF056266|AF056266 [3090915]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056265

Bacteriophage BF23 clone bf23.23.56r, genomic survey sequence
gi|3090914|gb|AF056265|AF056265 [3090914]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056264

Bacteriophage BF23 clone bf23.23.56f, genomic survey sequence
gi|3090913|gb|AF056264|AF056264 [3090913]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056263

Bacteriophage BF23 clone bf23.23.68f5r, genomic survey sequence
gi|3090912|gb|AF056263|AF056263 [3090912]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056262

Bacteriophage BF23 clone bf23.23.43fr.66f, genomic survey sequence
gi|3090911|gb|AF056262|AF056262 [3090911]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056261

Bacteriophage BF23 clone bf23.23.2fr, genomic survey sequence
gi|3090910|gb|AF056261|AF056261 [3090910]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056260

Bacteriophage BF23 clone bf23.23.55.f, genomic survey sequence
gi|3090909|gb|AF056260|AF056260 [3090909]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056259

Bacteriophage BF23 clone bf23.23.53.r, genomic survey sequence
gi|3090908|gb|AF056259|AF056259 [3090908]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

AF056258

Bacteriophage BF23 clone bf23.23.53.f, genomic survey sequence
gi|3090907|gb|AF056258|AF056258 [3090907]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056257

Bacteriophage BF23 clone bf23.23.52.r, genomic survey sequence
gi|3090906|gb|AF056257|AF056257 [3090906]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056256

Bacteriophage BF23 clone bf23.23.52.f, genomic survey sequence
gi|3090905|gb|AF056256|AF056256 [3090905]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056255

Bacteriophage BF23 clone bf23.23.49.r, genomic survey sequence
gi|3090904|gb|AF056255|AF056255 [3090904]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056254

Bacteriophage BF23 clone bf23.23.49.f, genomic survey sequence
gi|3090903|gb|AF056254|AF056254 [3090903]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056253

Bacteriophage BF23 clone bf23.23.48.r, genomic survey sequence
gi|3090902|gb|AF056253|AF056253 [3090902]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056252

Bacteriophage BF23 clone bf23.23.48.f, genomic survey sequence
gi|3090901|gb|AF056252|AF056252 [3090901]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056251

Bacteriophage BF23 clone bf23.23.44.r, genomic survey sequence
gi|3090900|gb|AF056251|AF056251 [3090900]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056250

Bacteriophage BF23 clone bf23.23.41.f, genomic survey sequence
gi|3090899|gb|AF056250|AF056250 [3090899]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056249

Bacteriophage BF23 clone bf23.23.22.a.r, genomic survey sequence
gi|3090898|gb|AF056249|AF056249 [3090898]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056248

Bacteriophage BF23 clone bf23.23.22.a.f, genomic survey sequence
gi|3090897|gb|AF056248|AF056248 [3090897]
(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

AF056247

Bacteriophage BF23 clone bf23.23.68.r, genomic survey sequence
gi|3090896|gb|AF056247|AF056247 [3090896]
(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

Z50114

Bacteriophage BF23 DNA for putative tail protein gene
gi|2464952|emb|Z50114|BF23LATE [2464952]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 protein link)

D12824

Bacteriophage BF23 genes for minor tail protein gp24 and major tail protein gp25, complete cds
gi|520578|dbj|D12824|BBF2TAIL [520578]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 3 nucleotide neighbors)

Z34953

Bacteriophage K3 ip9, ip7 and ip8 genes
gi|535261|emb|Z34953|MYK3IP978 [535261]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

Z35075

Bacteriophage K3 DNA for Ip3 and Ip4
gi|535229|emb|Z35075|MYEORF64K [535229]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

X05560

Bacteriophage K3 gene 38 for receptor recognizing protein
gi|15112|emb|X05560|MYK3G38 [15112]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

X04747

Bacteriophage K3 gene 37 for receptor recognizing protein
gi|15110|emb|X04747|MYK3G37 [15110]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

X01754

Bacteriophage K3 tail fiber gene 36
gi|15108|emb|X01754|MYK3F36 [15108]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

M16812

Bacteriophage K3 't' lysis gene, complete cds
gi|215503|gb|M16812|PK3LYST [215503]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

L46833

Bacteriophage K3 frd3, frd2 genes, complete cds
gi|951377|gb|L46833|PK3FRD32G [951377]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 protein links, or 2 nucleotide neighbors)

L43613

Bacteriophage K3 fibrin (wac) gene, complete cds
gi|903861|gb|L43613|PK3WAC [903861]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 4 nucleotide neighbors)

X01753

Bacteriophage Ox2 tail fiber gene 36

gi|15122|emb|X01753|MYOX2F36 [15122]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

L43612

Bacteriophage Ox2 fibrin (wac) gene, complete cds

gi|903848|gb|L43612|OX2WAC [903848]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 4 nucleotide neighbors)

Z46880

Bacteriophage OX2 stp gene

gi|599663|emb|Z46880|BPOX2STP [599663]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

X05675

Bacteriophage Ox2 gene 38 for receptor-recognizing protein and flanking regions

gi|15124|emb|X05675|MYOX2G38 [15124]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

M33533

Bacteriophage RB18 translational repressor protein (regA) and Orf43.1, complete cds

gi|216083|gb|M33533|RB18REGA [216083]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

AF033329

Bacteriophage RB18 single-stranded binding protein (gene 32) gene, partial cds, and 5' region

gi|2645788|gb|AF033329|AF033329 [2645788]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 11 nucleotide neighbors)

M86231

Bacteriophage RB69 gene 62, 3'end; RegA (regA) gene, complete cds

gi|215354|gb|M86231|P6962REGA [215354]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

AF033332

Bacteriophage RB69 single-stranded binding protein (gene 32) gene, partial cds, and 5' region

gi|2645794|gb|AF033332|AF033332 [2645794]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 12 nucleotide neighbors)

U34036

Bacteriophage RB69 DNA polymerase (43) gene, complete cds

gi|1237125|gb|U34036|BRU34036 [1237125]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

V01145

Bacteriophage H1 genome fragment Each Thymine given in this sequence represents a HMU-residue

(HMU = 5-hydroxymethyluracil)

gi|15557|emb|V01145|PODOH1 [15557]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

X05676

Bacteriophage M1 gene 38 for receptor recognizing protein and flanking regions

gi|15114|emb|X05676|MYM1G38 [15114]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

AF034575

Bacteriophage M1 putative integrase (int) gene, complete cds, and attP region, complete sequence

gi|2662472|gb|AF034575|AF034575 [2662472]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

AF033321

Bacteriophage M1 single-stranded binding protein (gene 32) gene, partial cds, and 5' region

gi|2645772|gb|AF033321|AF033321 [2645772]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 17 nucleotide neighbors)

X55190

Bacteriophage Tu1a 37 and 38 genes for receptor-recognizing proteins 37 and 38 (respectively), partial cds

gi|14860|emb|X55190|BPTUIA [14860]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

AF033334

Bacteriophage Tu1b single-stranded binding protein (gene 32) gene, partial cds, and 5' region

gi|2645798|gb|AF033334|AF033334 [2645798]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 5 nucleotide neighbors)

X55191

Bacteriophage Tu1b 37 gene for receptor-recognizing protein 37 (partial cds), 38 gene for receptor-recognizing protein 38, and t gene (partial cds)

gi|14863|emb|X55191|BPTUIB [14863]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 3 nucleotide neighbors)

X13065

Bacteriophage phi80 early region

gi|14800|emb|X13065|BP80ER [14800]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 8 protein links, or 6 nucleotide neighbors)

D00360

Bacteriophage phi80 cor gene

gi|217782|dbj|D00360|P8080COR [217782]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 protein link)

X01639

Bacteriophage phi 80 DNA-fragment with replication origin

gi|15828|emb|X01639|XCPHI80 [15828]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 25 nucleotide neighbors)

X04051

Lambdoid bacteriophage phi 80 int-xis region (integrase-excisionase region)

gi|15770|emb|X04051|STPHI80X [15770]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

X06751

Phage Phi80 DNA for major coat protein

gi|15768|emb|X06751|STPHI80C [15768]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 11 nucleotide neighbors)

X75949

Bacteriophage phi80 DNA for ORF x171.5 and ORF x171.28'

gi|458811|emb|X75949|ECORF171B [458811]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 28 nucleotide neighbors)

L40418

Bacteriophage phi-80 gene, complete cds

gi|1019107|gb|L40418|P80A [1019107]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

M24831

Bacteriophage phi-80 Tyr-tRNA gene, 3' end

gi|215363|gb|M24831|P80TGY [215363]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 43 nucleotide neighbors)

M10670

Bacteriophage phi-80 replication origin

gi|215361|gb|M10670|P80ORI [215361]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

M24825

Bacteriophage phi-80 RNA fragment

gi|215360|gb|M24825|P80M3A [215360]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 nucleotide neighbor)

M11919

Bacteriophage phi-80 cI immunity region encoding the N gene

gi|215358|gb|M11919|P80CI [215358]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

M10891

Bacteriophage phi-80 attP site DNA

gi|215357|gb|M10891|P80ATT [215357]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 nucleotide neighbor)

M19473

Bacteriophage 933J (from E.coli) proviral Shiga-like toxin type I subunits A and B genes, complete cds

gi|215072|gb|M19473|J93SLTI [215072]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 2 protein links, or 20 nucleotide neighbors)

Y10775

Bacteriophage 933W ileX, stx2A and stx2B genes

gi|1938206|emb|Y10775|BP933ILEX [1938206]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 36 nucleotide neighbors)

X83722

Bacteriophage 933W stx-IIIB gene

gi|1490229|emb|X83722|B933WSLT [1490229]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 20 nucleotide neighbors)

X07865

Bacteriophage 933W stx-II gene for Shiga-like toxin type II subunit A and B

gi|14892|emb|X07865|BWSLTII [14892]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 29 nucleotide neighbors)

M16625

Bacteriophage H19B (from E.coli) stxA and stxB genes encoding Shiga-like toxin I subunits A and B, complete cds

gi|215043|gb|M16625|H19BSLT [215043]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 24 nucleotide neighbors)

M17358

Bacteriophage H19B shiga-like toxin-I (SLT-I) A and B subunit DNA, complete cds
gi|215046|gb|M17358|H19BSLTA [215046]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 20 nucleotide neighbors)

U29728

Bacteriophage N4 single-stranded DNA-binding protein (N4SSB) gene, complete cds
gi|939708|gb|U29728|BNU29728 [939708]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, or 1 protein link)

J02580

Bacteriophage PA-2 (E.coli porcine strain isolate) Rz gene, 5'end; ORF2, outer membrane porin protein (lc) and ORF1 genes.
complete cds

gi|215366|gb|J02580|PA2LC [215366]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 4 nucleotide neighbors)

U32222

Bacteriophage 186, complete sequence

gi|3337249|gb|U32222|B1U32222 [3337249]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,6 MEDLINE links, 46 protein links, or 5 nucleotide neighbors)

X51522

Bacteriophage P4 complete DNA genome

gi|450916|emb|X51522|MYP4CG [450916]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 13 protein links, 6 nucleotide neighbors,
or 1 genome link)

X92588

Bacteriophage 82 orf33, orf151, orf56, orf96, rus, orf45, and Q genes

gi|1051111|emb|X92588|BAC82HOLL [1051111]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,7 protein links, or 1 nucleotide neighbor)

J02803

Bacteriophage 82 antitermination protein (Q) gene, complete cds

gi|215364|gb|J02803|P82Q [215364]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

U02466

Bacteriophage HK022 (cro), (cII) and (O) genes, complete cds, (P) gene, partial cds

gi|407285|gb|U02466|BHU02466 [407285]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 5 protein links, or 1 nucleotide neighbor)

M26291

Bacteriophage D108 regulatory DNA-binding protein (ner) gene, complete cds

gi|166194|gb|M26291|D18NER [166194]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

M11272

Bacteriophage D108 left-end DNA

gi|166193|gb|M11272|D18LEDNA [166193]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 2 nucleotide neighbors)

M18902

Bacteriophage D108 kil gene encoding a replication protein, 3' end; and containing three ORFs, complete cds

gi|166191|gb|M18902|D18KIL [166191]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

M10191

Bacteriophage D108, left end with Mu A protein binding sites L1 and L2

gi|166190|gb|M10191|D18BSL [166190]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 5 nucleotide neighbors)

J02447

bacteriophage d108 gene a 5' end

gi|166189|gb|J02447|D18AAA [166189]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

V00865

Bacteriophage D108 fragment from genes A and ner (C-terminus of ner and N-terminus of A)

gi|15437|emb|V00865|NCD108 [15437]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

X01914

Bacteriophage IKe gene for DNA binding protein

gi|14957|emb|X01914|INKEDBP [14957]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

AF064539

Bacteriophage N15, complete genome

gi|3192683|gb|AF064539|AF064539 [3192683]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 60 protein links, 26 nucleotide neighbors, or 1 genome link)

U02303

Bacteriophage If1, complete genome

gi|3676280|gb|U02303|B2U02303 [3676280]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 10 protein links, or 1 genome link)

AF007792

Bacteriophage Mu late morphogenetic region

gi|3551775|gb|AF007792|AF007792 [3551775]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 nucleotide neighbor)

U24159

Bacteriophage HP1 strain HP1c1, complete genome

gi|1046235|gb|U24159|BHU24159 [1046235]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 6 MEDLINE links, 41 protein links, 8 nucleotide neighbors, or 1 genome link)

Z71579

Bacteriophage S2 type A 5.6 kb DNA fragment

gi|1679806|emb|Z71579|BPHS1ADNA [1679806]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 3 MEDLINE links, 9 protein links, or 9 nucleotide neighbors)

X53238

Klebsiella sp. bacteriophage K11 gene 1 for RNA polymerase

gi|14984|emb|X53238|KSK11RPO [14984]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

X85010

Bacteriophage A511 ply511 gene

gi|853748|emb|X85010|BPA511PLY [853748]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link,3 protein links, or 1 nucleotide neighbor)

U29728

Bacteriophage N4 single-stranded DNA-binding protein (N4SSB) gene, complete cds

gi|939708|gb|U29728|BNU29728 [939708]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, or 1 protein link)

J02445

bacteriophage bo1 3'-terminal region rna

gi|166152|gb|J02445|BO1TR3 [166152]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 5 nucleotide neighbors)

L06183

Bacteriophage L5 (from *Leuconostoc oenos*) genome

gi|289353|gb|L06183|BL5GENM [289353]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 genome link)

AF074945

Mycoplasma arthritidis bacteriophage MAV1, complete genome

gi|3511243|gb|AF074945|AF074945 [3511243]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,15 protein links, 3 nucleotide neighbors, or 1 genome link)

L13696

Bacteriophage L2 (from *Mycoplasma*), complete genome

gi|289338|gb|L13696|BL2CG [289338]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 14 protein links, or 1 genome link)

X80191

Bacteriophage PP7 mRNA for maturation, coat, lysis and replicase proteins

gi|517237|emb|X80191|BPP7PR [517237]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 1 genome link)

M19377

Bacteriophage PF3 from *Pseudomonas aeruginosa* (New York strain), complete genome

gi|215380|gb|M19377|PF3COMNY [215380]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 9 protein links, or 5 nucleotide neighbors)

M11912

Bacteriophage PF3 from *Pseudomonas aeruginosa* (Nijmegen strain), complete genome

gi|215371|gb|M11912|PF3COMN [215371]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 9 protein links, 5 nucleotide neighbors, or 1 genome link)

V00605

Bacteriophage Pfl gene encoding DNA binding protein

gi|14970|emb|V00605|INOPF1 [14970]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 1 nucleotide neighbor)

L05626

Bacteriophage PR4 capsid protein (P6) gene, complete cds

gi|215735|gb|L05626|PR4P6MAJA [215735]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

D13409

Bacteriophage phiCTX (isolated from *Pseudomonas aeruginosa*) cosR, attP, int genes
gi|217776|dbj|D13409|BPHCOSR [217776]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 3 nucleotide neighbors)

D13408

Bacteriophage phiCTX (isolated from *Pseudomonas aeruginosa*) cosL, ctx genes
gi|217775|dbj|D13408|BPHCOSLCTX [217775]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, or 3 nucleotide neighbors)

M24832

Bacteriophage f2 coat protein gene, partial cds
gi|166228|gb|M24832|F2CRNACA [166228]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

S72011

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi|2618967|gb|AF017629|AF017629 [2618967]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017628

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi|2618964|gb|AF017628|AF017628 [2618964]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017627

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi|2618961|gb|AF017627|AF017627 [2618961]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017626

Bacteriophage 21 isocitrate dehydrogenase (icd) gene, partial cds; and integrase (int) gene, partial cds
gi|2618958|gb|AF017626|AF017626 [2618958]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 49 nucleotide neighbors)

AF017625

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi|2618955|gb|AF017625|AF017625 [2618955]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017624

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi|2618952|gb|AF017624|AF017624 [2618952]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017623

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi|2618949|gb|AF017623|AF017623 [2618949]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017622

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds
gi|2618946|gb|AF017622|AF017622 [2618946]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017621

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618943|gb|AF017621|AF017621 [2618943]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

D26449

Bacteriophage PS17 FI gene for tail sheath protein (gpFI) and FII gene for tail tube protein (gpFII), complete cds

gi|452162|dbj|D26449|BPSFIFII [452162]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 2 protein links)

X87627

Bacteriophage D3112 A and B genes

gi|974768|emb|X87627|BPD3112AB [974768]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

U32623

Bacteriophage D3 transcriptional activator CII (cII) gene, complete cds

gi|984852|gb|U32623|BDU32623 [984852]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 1 nucleotide neighbor)

L34781

Bacteriophage phi 11 holin homologue (ORF3) gene, complete cds and peptidoglycan hydrolase (lytA) gene, partial cds

gi|511838|gb|L34781|BPHHOLIN [511838]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 2 nucleotide neighbors)

L14810

Bacteriophage P22 (gp10) gene, complete cds, and (gp26) gene, complete cds

gi|294053|gb|L14810|P22GP1026X [294053]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

X87420

Bacteriophage ES18 genes 24, c2, cro, c1, 18, and oL and oR operators

gi|1143407|emb|X87420|BPES18GEN [1143407]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,5 protein links, or 9 nucleotide neighbors)

L42820

Bacteriophage BF23 tail protein (hrs) gene, complete cds

gi|1048680|gb|L42820|BBFHRS [1048680]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

X14980

Bacteriophage PRD1 XV gene for protein P15 (lytic enzyme)

gi|15802|emb|X14980|TEPRD1XV [15802]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

X06321

Bacteriophage PRD1 gene 8 for DNA terminal protein

gi|15800|emb|X06321|TEPRD18 [15800]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 10 nucleotide neighbors)

X14336

Filamentous Bacteriophage I2-2 genome

gi|14920|emb|X14336|NB122 [14920]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 9 protein links, 1 nucleotide neighbor, or 1 genome link)

L05001

Bacteriophage X glucosyl transferase gene, complete cds

gi|216044|gb|L05001|PXFCLUSYLT [216044]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

M29479

Bacteriophage p4 sid and psu genes partial cds, and delta gene, complete cds gi|215701|

gb|M29479|PP4SDP [215701]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 protein links, or 4 nucleotide neighbors)

SEG_PP4PSUSID

Bacteriophage P4 capsid size determination protein (sid) gene, 5' end

gi|215698|gb|SEG_PP4PSUSID [215698]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

M29650

Bacteriophage P4 polarity suppression protein (psu) gene, complete cds

gi|215697|gb|M29650|PP4PSUSID2 [215697]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M29651

Bacteriophage P4 capsid size determination protein (sid) gene, 5' end

gi|215696|gb|M29651|PP4PSUSID1 [215696]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M27748

Bacteriophage P4 gop, beta, and cII genes, complete cds and int gene, 3' end

gi|215691|gb|M27748|PP4GOPBC [215691]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 1 nucleotide neighbor)

K02750

Bacteriophage IKc, complete genome

gi|215061|gb|K02750|IKECG [215061]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 10 protein links, 4 nucleotide neighbors, or 1 genome link)

L40418

Bacteriophage phi-80 gene, complete cds

gi|1019107|gb|L40418|P80A [1019107]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

AF032122

Bacteriophage Sfil integrase (int) gene, partial cds; and bactoprenol glucosyl transferase (bgt), and glucosyl transferase II (gtrII) genes, complete cds

gi|2465412|gb|AF021347|AF021347 [2465412]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 2 nucleotide neighbors)

M35825

Bacteriophage SF6 fragment D lysozyme gene, complete cds

gi|216105|gb|M35825|SF6LYZ [216105]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 protein link)

Z35479

Bacteriophage C16 ipI gene

gi|534936|emb|Z35479|BC16IP1 [534936]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 2 nucleotide neighbors)

X12638

Bacteriophage 21 DNA for gene 2

gi|296141|emb|X12638|B21GENE2 [296141]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

X02501

Bacteriophage 21 DNA for left end sequence with genes 1 and 2

gi|15825|emb|X02501|XXPHA21 [15825]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 3 nucleotide neighbors)

M65239

Bacteriophage 21 lysis genes S, R, and Rz, complete cds

gi|215466|gb|M65239|PH2LYSGEN [215466]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

M58702

Bacteriophage 21 late gene regulatory region

gi|215465|gb|M58702|PH2LATEGE [215465]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

M81255

Bacteriophage 21 head gene operon

gi|215454|gb|M81255|PH2HEADTL [215454]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 10 protein links, or 4 nucleotide neighbors)

M23775

Bacteriophage 21 glycoprotein 1 gene, complete cds, and glycoprotein gene, 5' end

gi|215451|gb|M23775|PH2GPA [215451]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 3 nucleotide neighbors)

M61865

Bacteriophage 21 excisionase (xis), integrase (int) and isocitrate dehydrogenase (icd), complete cds

gi|215448|gb|M61865|PH22XISAA [215448]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 9 nucleotide neighbors)

S72011

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618967|gb|AF017629|AF017629 [2618967]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017628

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618964|gb|AF017628|AF017628 [2618964]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017627

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618961|gb|AF017627|AF017627 [2618961]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017626

Bacteriophage 21 isocitrate dehydrogenase (icd) gene, partial cds; and integrase (int) gene, partial cds

gi|2618958|gb|AF017626|AF017626 [2618958]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 49 nucleotide neighbors)

AF017625

242

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618955|gb|AF017625|AF017625 [2618955]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017624

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618952|gb|AF017624|AF017624 [2618952]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017623

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618949|gb|AF017623|AF017623 [2618949]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017622

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618946|gb|AF017622|AF017622 [2618946]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

AF017621

Bacteriophage 21 isocitrate dehydrogenase (icd) and integrase (int) genes, partial cds

gi|2618943|gb|AF017621|AF017621 [2618943]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 44 nucleotide neighbors)

M57455

Bacteriophage 42D (clone pDB17) (from *Staphylococcus aureus*) staphylokinase gene, complete cds

gi|215344|gb|M57455|P42STK [215344]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 9 nucleotide neighbors)

Y12633

Bacteriophage 85 DNA, promoter sequence of unknown gene

gi|2058285|emb|Y12633|B85PROM [2058285]

(View GenBank report, FASTA report, ASN.1 report, or Graphical view)

X98146

Bacteriophage P1 DNA sequence around the Op88 operator

gi|1359513|emb|X98146|BP1OP88OP [1359513]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 nucleotide neighbor)

Y07739

Staphylococcus phage Twort holTW, plyTW genes

gi|2764979|emb|Y07739|BPTWGHOLG [2764979]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 2 protein links)

L07580

Bacteriophage phi-11 rinA and rinB genes, required for the activation of Staphylococcal phage phi-11 int expression

gi|166160|gb|L07580|BPHRINAB [166160]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 2 protein links)

M34832

Bacteriophage phi-11 integrase (int) and excisionase (xis) genes, complete cds

gi|166157|gb|M34832|BPHINTXIS [166157]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

M20394

Bacteriophage phi-11 S.aureus attachment site (attP)

gi|166156|gb|M20394|BPHATTP [166156]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 4 nucleotide neighbors)

X23128

Bacteriophage phi-13 integrase gene

gi|758228|emb|X82312|PHI13INT [758228]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 protein link, or 3 nucleotide neighbors)

X61719

S.aureus phi-13 lysogen right chromosome/bacteriophage DNA junction

gi|46625|emb|X61719|SAP13RJNC [46625]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

X61718

S.aureus phi-13 lysogen left chromosomal/bacteriophage DNA junction

gi|46624|emb|X61718|SAP13LJNC [46624]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

X61717

Bacteriophage phi-13 core sequence for attachment

gi|14799|emb|X61717|BP13ATTTP [14799]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, or 3 nucleotide neighbors)

U01875

Bacteriophage phi-13 putative regulatory region and integrase (int) gene, partial cds

gi|437118|gb|U01875|U01875 [437118]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, or 4 nucleotide neighbors)

X67739

S.aureus Bacteriophage phi-42 attP gene

gi|14809|emb|X67739|BPATTPA [14809]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 3 nucleotide neighbors)

U01872

Bacteriophage phi-42 integrase (int) gene, complete cds

gi|437115|gb|U01872|U01872 [437115]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 2 protein links, or 3 nucleotide neighbors)

X94423

Staphylococcus aureus bacteriophage phi-42 DNA with ORFs (restriction modification system)

gi|1771597|emb|X94423|SARMS [1771597]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 1 nucleotide neighbor)

M27965

Bacteriophage L54a (from S.aureus) int and xis genes, complete cds

gi|215096|gb|M27965|L54INTXIS [215096]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, MEDLINE 1 link, 2 protein links, or 3 nucleotide neighbors)

U72397

Bacteriophage 80 alpha holin and amidase genes, complete cds

gi|1763241|gb|U72397|B8U72397 [1763241]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 2 nucleotide neighbors)

AB009866

Bacteriophage phi PVL proviral DNA, complete sequence

gi|3341907|dbj|AB009866|AB009866 [3341907]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,63 protein links, or 1 nucleotide neighbor)

Z47794

Bacteriophage Cp-1 DNA, complete genome

gi|2288892|emb|Z47794|BPCP1XX [2288892]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 28 protein links, 1 nucleotide neighbor, or 1 genome link)

SEG_CP7RSIT

Bacteriophage Cp-7 (S.pneumoniae) 5' inverted terminal repeat

gi|166186|gb|SEG_CP7RSIT [166186]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

M11635

Bacteriophage Cp-7 (S.pneumoniae) DNA, 3' inverted terminal repeat

gi|166185|gb|M11635|CP7RSIT2 [166185]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M11636

Bacteriophage Cp-7 (S.pneumoniae) 5' inverted terminal repeat

gi|166184|gb|M11636|CP7RSIT1 [166184]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

SEG_CP5RSIT

Bacteriophage Cp-5 (S.pneumoniae), 5' inverted terminal repeat

gi|166181|gb|SEG_CP5RSIT [166181]

(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 1 MEDLINE link)

M11633

Bacteriophage Cp-5 (S.pneumoniae) 3' inverted terminal repeat

gi|166180|gb|M11633|CP5RSIT2 [166180]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M11634

Bacteriophage Cp-5 (S.pneumoniae), 5' inverted terminal repeat

gi|166179|gb|M11634|CP5RSIT1 [166179]

(View GenBank report,FASTA report,ASN.1 report, or Graphical view)

M34780

Bacteriophage Cp-9 muramidase (cpl9) gene

gi|166187|gb|M34780|CP9CPL [166187]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 1 nucleotide neighbor)

M34652

Bacteriophage HB-3 amidase (hbl) gene, complete cds

gi|215055|gb|M34652|HB3HBLA [215055]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

U64984

Streptococcus pyogenes phage T12 repressor, excisionase (xis), integrase(int) and erythrogenic toxin A precursor (speA) genes, complete cds gi|1877426|gb|U40453|SPU40453 [1877426]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 4 protein links, or 22 nucleotide neighbors)

X12375

Phage CP-T1 (*Vibrio cholerae*) DNA for packaging signal (pac site)

gi|15435|emb|X12375|NCCPPAC [15435]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

AF087814

Vibrio cholerae filamentous bacteriophage fs-2 DNA, complete genome sequence

gi|3702207|dbj|AB002632|AB002632 [3702207]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 9 protein links, or 1 genome link)

D83518

Bacteriophage KVP40 gene for major capsid protein precursor, complete cds

gi|3046858|dbj|D83518|D83518 [3046858]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 1 protein link)

AF033322

Bacteriophage PST single-stranded binding protein (gene 32) gene, partial cds, and 5' region

gi|2645774|gb|AF033322|AF033322 [2645774]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 protein link, or 17 nucleotide neighbors)

X94331

Bacteriophage L cro, 24, c2, and c1 genes

gi|1469213|emb|X94331|BLCRO24C [1469213]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 protein links)

U82619

Shigella flexneri bacteriophage V glucosyl transferase (gtr), integrase (int) and excisionase (xis) genes, complete cds

gi|2465470|gb|U82619|SFU82619 [2465470]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 8 protein links, or 1 nucleotide neighbor)

246
Table 12

NCBI *Entrez* Nucleotide QUERY

Key words: bacteriophage and lysis

56 citations found (all selected)

AJ011581

Bacteriophage PS119 lysis genes 13, 19, 15, and packaging gene 3, complete cds
gi3676084|embl|AJ011581|BPS011581 [3676084]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 4 protein links, or 1 nucleotide neighbor)

AJ011580

Bacteriophage PS34 lysis genes 13, 19, 15, antiterminator gene 23, and packaging gene 3, complete cds
gi3676078|embl|AJ011580|BPS011580 [3676078]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 5 protein links, or 2 nucleotide neighbors)

AJ011579

Bacteriophage PS3 lysis genes 13, 19, 15, and packaging gene 3
gi3676073|embl|AJ011579|BPS011579 [3676073]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 4 protein links, or 1 nucleotide neighbor)

AF034975

Bacteriophage H-19B essential recombination function protein (erf), kil protein (kil), regulatory protein cIII (cIII), protein gp17 (17), N protein (N), cI protein (cI), cro protein (cro), cII protein (cII), O protein (O), P protein (P), ren protein (ren), Roi (roi), Q protein (Q), Shiga-like toxin A (slt-IA) and B (slt-IB) subunits, and putative holin protein (S) genes, complete cds
gi2668751|gb|AF034975 [2668751]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 20 protein links, or 30 nucleotide neighbors)

U37314

Bacteriophage lambda Rz1 protein precursor (Rz1) gene, complete cds
gi1017780|gb|U37314|BLU37314 [1017780]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 1 protein link, or 9 nucleotide neighbors)

U00005

E. coli hflA locus encoding the hflX, hflK and hflC genes, hflq gene, complete cds; miaA gene, partial cds
gi436153|gb|U00005|ECHOFLA [436153]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 4 MEDLINE links, 1 protein link, or 9 nucleotide neighbors)

links, 5 protein links, or 8 nucleotide neighbors)

U32222

Bacteriophage 186, complete sequence

gi|3337249|gb|U32222|B1U32222 [3337249]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 6 MEDLINE links, 46 protein links, or 5 nucleotide neighbors)

AF064539

Bacteriophage N15, complete genome

gi|3192683|gb|AF064539|AF064539 [3192683]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 60 protein links, 26 nucleotide neighbors, or 1 genome link)

AF063097

Bacteriophage P2, complete genome

gi|3139086|gb|AF063097|AF063097 [3139086]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 21 MEDLINE links, 42 protein links, 3 nucleotide neighbors, or 1 genome link)

Z97974

Bacteriophage phiadh lys, hol, intG, rad, and tec genes

gi|2707950|emb|Z97974|BPHIADH [2707950]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 9 protein links, or 1 nucleotide neighbor)

AF059243

Bacteriophage NL95, complete genome

gi|3088545|gb|AF059243|AF059243 [3088545]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 4 protein links, 3 nucleotide neighbors, or 1 genome link)

AF052431

Bacteriophage M11 A-protein, coat protein, A1-protein, and replicase genes, complete cds

gi|2981208|gb|AF052431| [2981208]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 4 protein links, or 8 nucleotide neighbors)

Y07739

Staphylococcus phage Twort holTW, plyTW genes

gi|2764979|emb|Y07739|BPTWGHOLG [2764979]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 2 protein links)

X94331

Bacteriophage L cro, 24, c2, and c1 genes
gil1469213|emb|X94331|BLCRO24C [1469213]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 4 protein links)

X78410

Bacteriophage phiadh holin and lysin genes
gil793848|emb|X78410|LGHOLLYS [793848]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

X99260

Bacteriophage B103 genomic sequence
gil1429229|emb|X99260|BB103G [1429229]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 17 protein links, or 12 nucleotide neighbors)

AJ000741

Bacteriophage P1 darA operon
gil2462938|emb|AJ000741|BPAP7641 [2462938]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 10 protein links, or 31 nucleotide neighbors)

X87420

Bacteriophage ES18 genes 24, c2, cro, c1, 18, and oL and oR operators
gil1143407|emb|X87420|BPES18GEN [1143407]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 5 protein links, or 9 nucleotide neighbors)

L35561

Bacteriophage phi-105 ORFs 1-3
gil532218|gb|L35561|PH5ORFHTR [532218]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 3 protein links)

D10027

Group II RNA coliphage GA genome
gil217784|dbj|D10027|PGAXX [217784]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, 5 nucleotide neighbors, or 1 genome link)

V01128

Bacteriophage phi-X174 (cs70 mutation) complete genome
gil15535|emb|V01128|PHIX174 [15535]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 4 MEDLINE links, 11 protein links, or 26 nucleotide neighbors)

S81763

coat gene...replicase gene (bacteriophage KU1, host=Escherichia coli, group II RNA phage. Genomic RNA, 3 genes, 120 nt)
gil1438766|gb|S81763|S81763 [1438766]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 1 MEDLINE link)

U38906

Bacteriophage r1t integrase, repressor protein (rro), dUTPase, holin and lysin genes, complete cds
gil1353517|gb|U38906|BRU38906 [1353517]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 50 protein links, or 3 nucleotide neighbors)

X91149

Bacteriophage phi-C31 DNA cos region
gil1107473|emb|X91149|APHIC31C [1107473]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 6 protein links, or 1 nucleotide neighbor)

V00642

phage MS2 genome
gil15081|emb|V00642|LEMS2X [15081]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 8 MEDLINE links, 4 protein links, or 20 nucleotide neighbors)

V01146

Genome of bacteriophage T7
gil431187|emb|V01146|T7CG [431187]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 13 MEDLINE links, 60 protein links, 105 nucleotide neighbors, or 1 genome link)

X78401

Bacteriophage P22 right operon, orf 48, replication genes 18 and 12, nin region genes, ninG phosphatase, late control gene 23, orf 60, complete cds, late control region, start of lysis gene 13
gil512343|emb|X78401|POP22NIN [512343]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 13 protein links, or 4 nucleotide neighbors)

Y00408

Bacteriophage T4 gene t for lysis protein
gil15368|emb|Y00408|MYT4T [15368]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

Z26590

Bacteriophage mv4 lysA and lysB genes
gi1410500|emb1Z26590|MV4LYSAB [410500]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 4 protein links)

X07809

Phage phiX174 lysis (E) gene upstream region
gi115094|emb1X07809|M1PHLXE [15094]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 4 nucleotide neighbors)

Z34528

Lactococcal bacteriophage c2 lysin gene
gi1506455|emb1Z34528|LBC2LYSIN [506455]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 4 nucleotide neighbors)

X15031

Bacteriophage fr RNA genome
gi115071|emb1X15031|LEBFRX [15071]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 4 protein links, 9 nucleotide neighbors, or 1 genome link)

X80191

Bacteriophage PP7 mRNA for maturation, coat, lysis and replicase proteins
gi1517237|emb1X80191|BPP7PR [517237]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 4 protein links, or 1 genome link)

X85010

Bacteriophage A511 ply511 gene
gi1853748|emb1X85010|BPA511PLY [853748]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

X85009

Bacteriophage A500 hol500 and ply500 genes
gi1853744|emb1X85009|BPA500PLY [853744]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 4 nucleotide neighbors)

X85008

Bacteriophage A118 hol118 and ply118 genes
gi1853740|emb1X85008|BPA118PLY [853740]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

251

Z35638

Bacteriophage phi-X174 genes for lysis protein and beta-lactamase
gil520996|embl|Z35638|BPLYSPR [520996]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 2 protein links, or 516 nucleotide neighbors)

J02459

Bacteriophage lambda, complete genome
gil215104|gb|J02459|LAMCG [215104]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 87 MEDLINE
links, 67 protein links, 190 nucleotide neighbors, or 1 genome link)

X87674

Bacteriophage P1 lydA & lydB genes
gil974763|embl|X87674|BACP1LYD [974763]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 2 protein links, or 2 nucleotide neighbors)

X87673

Bacteriophage P1 gene 17
gil974761|embl|X87673|BACP117 [974761]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 1 protein link, or 1 nucleotide neighbor)

M14784

Bacteriophage T3 strain amNG220B right end, tail fiber protein, lysis
protein and DNA packaging proteins, complete cds
gil215810|gb|M14784|PT3RE [215810]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 9 protein links, or 10 nucleotide neighbors)

M11813

Bacteriophage PZA (from B.subtilis), complete genome
gil216046|gb|M11813|PZACG [216046]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 3 MEDLINE
links, 27 protein links, 17 nucleotide neighbors, or 1 genome link)

M16812

Bacteriophage K3 't' lysis gene, complete cds
gil215503|gb|M16812|PK3LYST [215503]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE
link, 1 protein link, or 4 nucleotide neighbors)

J04356

Bacteriophage P22 proteins 15 (complete cds), and 19 (3' end) genes
gil215265|gb|J04356|P2215P [215265]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 2 nucleotide neighbors)

J04343

Bacteriophage JP34 coat and lysis protein genes, complete cds, and replicase protein gene, 5' end
gi|215076|gb|J04343|JP3COLY [215076]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 2 nucleotide neighbors)

J02482

Bacteriophage phi-X174, complete genome
gi|216019|gb|J02482|PX1CG [216019]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,23 MEDLINE links, 11 protein links, 26 nucleotide neighbors, or 1 genome link)

M99441

Bacteriophage T4 anti-sigma 70 protein (asiA) gene, complete cds and lysis protein, 3' end
gi|215820|gb|M99441|PT4ASIA [215820]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE links, 2 protein links, or 2 nucleotide neighbors)

M65239

Bacteriophage 21 lysis genes S, R, and Rz, complete cds
gi|215466|gb|M65239|PH2LYSGEN [215466]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

M10637

Phage G4 D/E overlapping gene system, encoding D (morphogenetic) and E (lysis) proteins
gi|215427|gb|M10637|PG4DE [215427]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 12 nucleotide neighbors)

J02454

Bacteriophage G4, complete genome
gi|215415|gb|J02454|PG4CG [215415]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,6 MEDLINE links, 11 protein links, 20 nucleotide neighbors, or 1 genome link)

J02580

Bacteriophage PA-2 (E.coli porcine strain isolate) Rz gene, 5' end; ORF2, outer membrane porin protein (Ic) and ORF1 genes, complete cds
gi|215366|gb|J02580|PA2LC [215366]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 4 protein links, or 4 nucleotide neighbors)

M14782

Bacillus phage phi-29 head morphogenesis, major head protein, head fiber protein, tail protein, upper collar protein, lower collar protein, pre-neck appendage protein, morphogenesis(13), lysis, morphogenesis(15), encapsidation genes, complete cds
gi|2153231|gb|M14782|P29LATE2 [215323]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 11 protein links, or 11 nucleotide neighbors)

M10997

Bacteriophage P22 lysis genes 13 and 19, complete cds
gi|2152621|gb|M10997|P221319 [215262]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 3 nucleotide neighbors)

J02467

Bacteriophage MS2, complete genome
gi|2152321|gb|J02467|MS2CG [215232]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,8 MEDLINE links, 4 protein links, 20 nucleotide neighbors, or 1 genome link)

M14035

Bacteriophage lambda lysis S gene with mutations leading to nonlethality of S in the plasmid pRG1
gi|2151801|gb|M14035|LAMLYS [215180]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 1 protein link, or 14 nucleotide neighbors)

U04309

Bacteriophage phi-LC3 putative holin (lysA) gene and putative murein hydrolase (lysB) gene, complete cds
gi|5307961|gb|U04309|BPU04309 [530796]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

Table 13

NCBI Entrez Nucleotide QUERY**Key word: holin****51 citations found (all selected)****AF034975**

Bacteriophage H-19B essential recombination function protein (erf), kil protein (kil), regulatory protein cIII (cIII), protein gp17 (17), N protein (N), cI protein (cI), cro protein (cro), cII protein (cII), O protein (O), P protein (P), ren protein (ren), Roi (roi), Q protein (Q), Shiga-like toxin A (slt-IA) and B (slt-IB) subunits, and putative holin protein (S) genes, complete cds
 gi|2668751|gb|AF034975| [2668751]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 20 protein links, or 30 nucleotide neighbors)

U52961

Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds
 gi|1841516|gb|U52961|SAU52961 [1841516]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

U28154

Haemophilus somnus cryptic prophage genes, capsid scaffolding protein gene, partial cds, major capsid protein precursor, endonuclease, capsid completion protein, tail synthesis proteins, holin, and lysozyme genes, complete cds
 gi|1765928|gb|U28154|HSU28154 [1765928]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, or 13 protein links)

AF032122

Streptococcus thermophilus bacteriophage Sfi19 central region of genome
 gi|2935682|gb|AF032122| [2935682]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 14 protein links, or 2 nucleotide neighbors)

AF032121

Streptococcus thermophilus bacteriophage Sfi21 central region of genome
 gi|2935667|gb|AF032121|AF032121 [2935667]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 14 protein links, or 2 nucleotide neighbors)

AF021803

Bacillus subtilis 168 prophage SPbeta N-acetylmuramoyl-L-alanine amidase (blyA), holin-like protein (bh1A), holin-like protein (bh1B), and yolK genes, complete cds; and yolJ gene, partial cds
gi2997594|gb|AF021803|AF021803 [2997594]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 5 protein links, or 1 nucleotide neighbor)

AF057033

Streptococcus thermophilus bacteriophage sfil1 gp502 (orf502), gp284 (orf284), gp129 (orf129), gp193 (orf193), gp119 (orf119), gp348 (orf348), gp53 (orf53), gp113 (orf113), gp104 (orf104), gp114 (orf114), gp128 (orf128), gp168 (orf168), gp117 (orf117), gp105 (orf105), putative minor tail protein (orf1510), putative minor structural protein (orf512), putative minor structural protein (orf1000), gp373 (orf373), gp57 (orf57), putative anti-receptor (orf695), putative minor structural protein (orf669), gp149 (orf149), putative holin (orf141), putative holin (orf87), and lysin (orf288) genes, complete cds
gi3320432|gb|AF057033|AF057033 [3320432]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,25 protein links, or 1 nucleotide neighbor)

U32222

Bacteriophage 186, complete sequence
gi3337249|gb|U32222|B1U32222 [3337249]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,6 MEDLINE links, 46 protein links, or 5 nucleotide neighbors)

AB009866

Bacteriophage phi PVL proviral DNA, complete sequence
gi3341907|dbj|AB009866|AB009866 [3341907]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,63 protein links, or 1 nucleotide neighbor)

AF009630

Bacteriophage bIL170, complete genome
gi3282260|gb|AF009630|AF009630 [3282260]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,63 protein links, 3 nucleotide neighbors, or 1 genome link)

AF064539

Bacteriophage N15, complete genome

gi|3192683|gb|AF064539|AF064539 [3192683]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE
 links, 60 protein links, 26 nucleotide neighbors, or 1 genome link)

AF063097

Bacteriophage P2, complete genome
 gi|3139086|gb|AF063097|AF063097 [3139086]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 21 MEDLINE
 links, 42 protein links, 3 nucleotide neighbors, or 1 genome link)

Z97974

Bacteriophage phiadh lys, hol, intG, rad, and tec genes
 gi|2707950|emb|Z97974|BPHLADH [2707950]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE
 links, 9 protein links, or 1 nucleotide neighbor)

X95646

Streptococcus thermophilus bacteriophage Sfi21 DNA; lysogeny module,
 8141 bp
 gi|2292747|emb|X95646|BSFI21LYS [2292747]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE
 links, 19 protein links, or 3 nucleotide neighbors)

SEG_LLHLYSINO

Bacteriophage LL-H structural protein gene, partial cds; minor
 structural protein gp61 (g57), unknown protein, unknown protein,
 structural protein (g20), unknown protein, unknown protein, major capsid
 protein (g34), main tail protein gp19 (g17), holin (hol), muramidase
 (mur), unknown protein, unknown protein, unknown protein, unknown
 protein, unknown protein, and unknown protein genes, complete cds;
 unknown protein gene, partial cds; and unknown protein, unknown protein,
 unknown protein, unknown protein, unknown protein, minor structural
 protein gp75 (g70), minor structural protein gp89 (g88), minor
 structural protein gp58 (g71), unknown protein, unknown protein, unknown
 protein, and unknown protein genes, complete cds
 gi|1004337|gb|SEG_LLHLYSINO [1004337]
 (View GenBank report, FASTA report, ASN.1 report, Graphical view, 4 MEDLINE
 links, 31 protein links, or 1 nucleotide neighbor)

M96254

Bacteriophage LL-H holin (hol), muramidase (mur), and unknown protein
 genes, complete cds
 gi|1004336|gb|M96254|LLHLYSINO3 [1004336]
 (View GenBank report, FASTA report, ASN.1 report, or Graphical view)

Y07740

Staphylococcus phage 187 ply187 and hol187 genes
gil2764982|embl|Y07740|BP187PLYH [2764982]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, or 2 protein links)

U88974

Streptococcus thermophilus bacteriophage 01205 DNA sequence
gil2444080|gb|U88974| [2444080]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 57 protein links, or 6 nucleotide neighbors)

Z99117

Bacillus subtilis complete genome (section 14 of 21): from 2599451 to 2812870
gil2634966|embl|Z99117|BSUB0014 [2634966]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 233 protein links, 51 nucleotide neighbors, or 1 genome link)

Z99115

Bacillus subtilis complete genome (section 12 of 21): from 2195541 to 2409220
gil2634478|embl|Z99115|BSUB0012 [2634478]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 244 protein links, 64 nucleotide neighbors, or 1 genome link)

Z99110

Bacillus subtilis complete genome (section 7 of 21): from 1194391 to 1411140
gil2633472|embl|Z99110|BSUB0007 [2633472]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 226 protein links, 31 nucleotide neighbors, or 1 genome link)

X78410

Bacteriophage phiadh holin and lysin genes
gil793848|embl|X78410|LGHOLLYS [793848]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

Z93946

258

Bacteriophage Dp-1 dph and pal genes and 5 open reading frames
gil1934760|embl|Z93946|BPDP1ORFS [1934760]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 6
protein links)

AF011378

Bacteriophage sk1 complete genome
gil2392824|gb|AF011378|AF011378 [2392824]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,54 protein
links, 2 nucleotide neighbors, or 1 genome link)

Z47794

Bacteriophage Cp-1 DNA, complete genome
gil2288892|embl|Z47794|BPDP1XX [2288892]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,3 MEDLINE
links, 28 protein links, 1 nucleotide neighbor, or 1 genome link)

L35561

Bacteriophage phi-105 ORFs 1-3
gil532218|gb|L35561|PH5ORFHTR [532218]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, or 3 protein links)

D49712

Bacillus licheniformis DNA for ORFs, xpaL2 homologous protein and xpaL1
homologous protein, complete and partial cds
gil1514423|dbj|D49712|D49712 [1514423]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE
links, or 4 protein links)

X90511

Lactobacillus bacteriophage phig1e DNA for Rorf162, Holin, Lysin, and
Rorf175 genes
gil1926386|embl|X90511|LBPH1HOL [1926386]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,4 protein
links, or 1 nucleotide neighbor)

X98106

Lactobacillus bacteriophage phig1e complete genomic DNA
gil1926320|embl|X98106|LBPH1G1E [1926320]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE

link, 50 protein links, or 4 nucleotide neighbors²⁵⁹)

U72397

Bacteriophage 80 alpha holin and amidase genes, complete cds
gil1763241|gb|U72397|B8U72397 [1763241]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 protein links, or 2 nucleotide neighbors)

U38906

Bacteriophage rlt integrase, repressor protein (rro), dUTPase, holin and lysin genes, complete cds
gil1353517|gb|U38906|BRU38906 [1353517]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,2 MEDLINE links, 50 protein links, or 3 nucleotide neighbors)

X91149

Bacteriophage phi-C31 DNA cos region
gil1107473|emb|X91149|APHIC31C [1107473]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 6 protein links, or 1 nucleotide neighbor)

U24159

Bacteriophage HP1 strain HP1c1, complete genome
gil1046235|gb|U24159|BHU24159 [1046235]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,6 MEDLINE links, 41 protein links, 8 nucleotide neighbors, or 1 genome link)

Z26590

Bacteriophage mv4 lysA and lysB genes
gil410500|emb|Z26590|MV4LYSAB [410500]
(View GenBank report,FASTA report,ASN.1 report,Graphical view, or 4 protein links)

Z70177

B.subtilis-DNA (28 kb PBSX/skin element region)
gil1225934|emb|Z70177|BSPBSXSE [1225934]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,32 protein links, or 4 nucleotide neighbors)

Z36941

260

B.subtilis defective prophage PBSX xhlA, xhlB, and xylA genes
gil535793|embl|Z36941|BSPBSXXHL [535793]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,4 protein
links, or 5 nucleotide neighbors)

X89234

L.innocua DNA for phagelysin and holin gene
gil1134844|embl|X89234|LICPLYHOL [1134844]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 2 protein links, or 4 nucleotide neighbors)

X85010

Bacteriophage A511 ply511 gene
gil853748|embl|X85010|BPA511PLY [853748]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 3 protein links, or 1 nucleotide neighbor)

X85009

Bacteriophage A500 hol500 and ply500 genes
gil853744|embl|X85009|BPA500PLY [853744]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 3 protein links, or 4 nucleotide neighbors)

X85008

Bacteriophage A118 hol118 and ply118 genes
gil853740|embl|X85008|BPA118PLY [853740]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 3 protein links, or 1 nucleotide neighbor)

L34781

Bacteriophage phi 11 holin homologue (ORF3) gene, complete cds and
peptidoglycan hydrolase (lytA) gene, partial cds
gil511838|gb|L34781|BPHHOLIN [511838]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE
link, 4 protein links, or 2 nucleotide neighbors)

U11698

Serratia marcescens SM6 extracellular secretory protein (nucE), putative
phage lysozyme (nucD), and transcriptional activator (nucC) genes,
complete cds
gil509550|gb|U11698|SMU11698 [509550]
(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE

link, 3 protein links, or 1 nucleotide neighbor)

U31763

Serratia marcescens phage-holin analog protein (regA), putative phage lysozyme (regB), and transcriptional activator (regC) genes, complete cds

gi1965068|gb|U31763|SMU31763 [965068]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 3 protein links, or 1 nucleotide neighbor)

X87674

Bacteriophage P1 lydA & lydB genes

gi1974763|emb|X87674|BACP1LYD [974763]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 2 protein links, or 2 nucleotide neighbors)

L48605

Bacteriophage c2 complete genome

gi1146276|gb|L48605|C2PVC [1146276]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 3 MEDLINE links, 39 protein links, 3 nucleotide neighbors, or 1 genome link)

L33769

Bacteriophage bIL67 DNA polymerase subunit (ORF3-5), essential recombination protein (ORF13), lysin (ORF24), minor tail protein (ORF31), terminase subunit (ORF32), holin (ORF37), unknown protein (ORF 1-2, 6-12, 14-23, 25-30, 33-36), complete genome

gi1522252|gb|L33769|L67CG [522252]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 37 protein links, 2 nucleotide neighbors, or 1 genome link)

L31348

Bacteriophage Tuc2009 integrase (int) gene, complete cds; lysin (lys) gene, 3' end

gi1508612|gb|L31348|TU2INT [508612]

(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 3 protein links, or 3 nucleotide neighbors)

L31364

Bacteriophage Tuc2009 holin (S) gene, complete cds; lysin (lys) gene, complete cds

gi1496281|gb|L31364|TU2SLYS [496281]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

L31366

Bacteriophage Tuc2009 structural protein (mp2) gene, complete cds
gil496278|gb|L31366|TU2MP2A [496278]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

L31365

Bacteriophage Tuc2009 structural protein (mp1) gene, complete cds
gil496276|gb|L31365|TU2MP1A [496276]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, or 1 protein link)

U04309

Bacteriophage phi-LC3 putative holin (lysA) gene and putative murein
hydrolase (lysB) gene, complete cds
gil530796|gb|U04309|BPU04309 [530796]

(View GenBank report,FASTA report,ASN.1 report,Graphical view,1 MEDLINE link, 2 protein links, or 1 nucleotide neighbor)

Table 14

NCBI *Entrez* Nucleotide QUERY**Key word: bacteriophage and kil****5 citations found (all selected)****AF034975**

Bacteriophage H-19B essential recombination function protein (erf), kil protein (kil), regulatory protein cIII (cIII), protein gp17 (17), N protein (N), cI protein (cI), cro protein (cro), cII protein (cII), O protein (O), P protein (P), ren protein (ren), Roi (roi), Q protein (Q), Shiga-like toxin A (slt-IA) and B (slt-IB) subunits, and putative holin protein (S) genes, complete cds
gil2668751|gb|AF034975| [2668751]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 20 protein links, or 30 nucleotide neighbors)

X15637

Bacteriophage P22 P(L) operon encompassing *ral*, *17*, *kil* and *arf* genes
gil15646|emb|X15637|POP22PL [15646]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 7 protein links, or 2 nucleotide neighbors)

J02459

Bacteriophage lambda, complete genome
gil215104|gb|J02459|LAMCG [215104]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 87 MEDLINE links, 67 protein links, 190 nucleotide neighbors, or 1 genome link)

M64097

Bacteriophage Mu left end
gil215543|gb|M64097|PMULEFTEN [215543]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 2 MEDLINE links, 39 protein links, or 15 nucleotide neighbors)

M18902

Bacteriophage D108 kil gene encoding a replication protein, 3' end; and containing three ORFs, complete cds
gil166191|gb|M18902|D18KIL [166191]
(View GenBank report, FASTA report, ASN.1 report, Graphical view, 1 MEDLINE link, 1 protein link, or 3 nucleotide neighbors)

Table 15

U77328	V01282	U11787	U93688	A47599	D21131	U76864	U38428
AF151117	AF121672	U11786	U93687	A47598	D30690	U76863	U66665
AF151218	AF072726	U11785	AJ224764	A47597	D14711	U76862	U66664
AF146368	AF115379	U11784	AF064774	A47596	D90119	U76861	U66663
AF144661	AF034153	U11783	AF064773	A47595	D00730	U76860	X87104
AF132117	AF029244	U11782	Y14370	A47594	D83357	U76859	X87105
Y15477	U67965	U11781	AF065394	A44534	D83356	U76858	X89233
Y09928	U96610	U11780	AF062376	A44533	D83355	U76857	M28521
Y09594	U96609	U11779	AF062375	A44529	D83354	U76855	U54636
AF134905	U73027	U11778	AF062374	A44528	D83353	U76854	U46541
AB019536	U73026	U11777	AF062373	A44527	D12572	U76853	L14017
AJ237696	U73025	U11776	AB007500	A44526	D86727	U76852	U60589
AF106851	AF068904	U11775	Y09924	A44525	D86240	U76851	X48003
AF106850	U60050	U11774	U63529	A39696	D67075	U76850	M37889
AF106849	D10907	U11773	AF033191	AF001783	D67074	U76849	V01281
M26321	D10906	AF053772	Y15856	AF001782	U97062	U76848	X97985
AF060191	AF053140	AF053771	AB000439	L77194	U96620	U76847	X00127
AF060190	AB013298	AF029731	AF041467	AF003593	U96619	Y09929	X03286
AF060189	Y16431	AF027155	Y14051	AF003592	Z84573	Y09570	X62282
AF060188	AF076684	AF024571	U82085	X73889	AB001896	X95848	X01645
AF060187	AF076683	U87144	AF026122	X74219	Y07645	Y09428	X16471
AF060186	Y13225	AF086644	AF026121	Y10419	U92441	S76611	X52734
AF060185	AF094826	AJ223781	AF026120	M63177	U91741	S76213	X13290
AF060184	AJ223480	AF076030	AB009635	E08773	U29454	S75707	X66088
AF036324	AF093548	AF044951	AB006796	E07163	U29478	S75706	Z30588
AF036323	AJ005352	AF044906	U39769	E07162	U77374	S75705	X16457
AF053568	AF051916	AF044905	D00184	E07161	L42945	S76270	X00342
AJ132841	Y09927	AF044904	X56628	E07160	U38429	S72497	V01287
Y13766	AF051917	AF044903	AF033018	E07159	U81980	S72488	X61307
AF101234	S77058	AF044902	AF034076	E07158	X55185	S74031	Y00356
AJ133520	S65052	AF044901	D82063	E07157	V01278	S67449	X06603
AJ133495	AF009671	AF044900	D76414	E07156	U31979	U75367	Z93205
AJ132803	U81973	AF044899	U57060	E07155	X91786	U75368	X64172
AB016487	U77308	AF044898	D89066	E03836	U36912	U31175	X72700
AB016431	U20869	AF044897	U85095	E03835	U36911	X53096	X60827
AB015981	U89396	AF044075	U85097	E03526	U36910	X53951	X64389
AB015195	U94706	AF044074	U85096	E02873	U64885	X53952	X62288
AF107307	U41072	AF044073	D42078	E01690	U76872	X03408	X55798
AF079518	U52961	AF044072	AF015929	E00876	U76871	U50629	X58434
AJ223806	U21636	AF044071	D10369	E00203	U76870	U38656	X06627
Y18018	U65000	AF044070	A48955	D83951	U76869	U58139	X12831
Y17795	U48826	AF044069	A48501	D17366	U76868	A31894	X07371
AJ005647	U20503	AF044068	A48500	D42144	U76867	L42943	X02529
AJ005646	U11789	AF044067	A48499	D42143	U76866	U51474	Y00688
AJ005645	U11788	AF044066	A47600	D10489	U76865	U50077	X04121
X59477	X54338	A12915	U51133	M63176	M10500	L01055	M63917
X59478	X51661	A12913	U51132	L11998	M10499	M83994	M58515
X63598	X05815	A12906	X02588	L05004	AH000934	J03947	L10909
X52593	X15574	A12905	X61716	L42764	M10498	J03479	M15067

X76490	Y07536	A12904	X61719	M32103	M10497	M64724	M92376
X81586	X02166	A12903	X61718	U10927	M18264	M14372	M62650
X72014	Z49245	A12902	X67743	AH003057	J01786	M14371	M32312
X72013	X16298	A12901	X67742	M73535	M33833	M14374	M20393
X71437	Z18852	A12900	X67741	M73536	M32470	M15215	M90536
X62992	X68417	A12899	X67740	U20782	M20270	M36694	M21854
X52594	X68425	A12898	X67738	L37598	J03323	M37915	M36771
X14827	X17679	A12897	U02910	L37597	M33479	M12715	L14020
X13404	X63072	A12896	AH003349	L36472	M94061	J04151	M81736
X17301	X02872	A09523	M11118	L25288	M37888	L22566	U11702
X17688	V01277	A04518	M18086	L25893	M76714	L13379	L19300
X03097	X52543	A04517	U19459	K02687	M17123	L13378	L25372
Z16422	A19943	A04512	U35773	L23109	M97169	L13377	L22565
Z33409	A19942	L41499	U26702	L07778	M81346	L13376	M58516
Z33408	A19941	U19770	U21221	M90056	M90693	L13375	U06462
Z33407	A19940	X53818	U36379	J02615	M25257	L13374	L19298
Z33406	A19939	M20129	U06451	M18970	M25256	M17348	M80252
Z33405	A19938	L43098	U35036	K02985	M25255	M17357	L11530
Z33404	A19937	L43082	U20794	M21136	M25254	M17347	
X75439	A19936	X03216	L25426	M10501	M25253	M28364	
X62587	A17958	X70648	M86227	AH000935	M25252	M21319	

Table 16

Phage 44AHJD complete genome sequence. 16668 nucleotides.

```

1      tccatttctt tactaaactt aaaaatgctg tgcaacaact taaccaactt atctaacctt ttacatattc
71     atcaaataca aaatttatgt atctattgac ttttattcaa aattatgatt tcaacatata ataaaattaa
141    tttacttatt taatatattc atgatataat tagttataaa atatttggag gtgtataaat gacagaattt
211    gatgaatcg taaaaccaga cgacaaagaa gaaacttcag aatcaactga agaaaattta gaatcaactg
281    aagaaacttc agaatcaact gaagaatcaa ctgaagaatc aactgaagaa tcaactgaag ataaaacagt
351    agaaacaatc gaagaagaaa atgaaaacaa attagaacct actacaacag atgaagatag ttcgaaattt
421    gaccctgttg tattagaaca acgtattgct tcattagaac aacaagtgc tactttttta tcttcacaaa
491    tgcaacaacc acaacaagta caacaaacac aatcagatgt aacagaatca aacaaagaag ataacgacta
561    ttcagatgaa gaactagttg ataagttaga tttagattag gaggaattta aacatgtatg agggaaacaa
631    catgcgttct atgatgggta catcatatga agattcaaga ttaataaatac gaacagaatt aaatgaaaac
701    atgtcaattg atacaataaa aagtgaagat agttatgggtg tacaattcca ttcactttca aaacaatcat
771    ttacagggtg cgttgaggag gaataataaa ttatggcaca acaatctaca aaaaatgaaa ctgcactttt
841    agtagcaaaag tcagctaaat cagcgttaca agattttaat catgattatt caaaatcttg gacatttggc
911    gacaaatggg ataattcaaa tacaatgttc gaacatttg taaataaata tttattccct aagatttaag
981    agacttttatt aatcgatatt gcattaggta atcgtttaa ttggttagct aaagagcaag attttatgg
1051   acaatatagt gaagaatcag tgattatgga cacagtacca attaacatgg acttatctaa aaatgaggaa
1121   ttaatgttga aacgtaatga tccacgtatg gcaactaagt tatatggtaa cggaaatttg aagaaacaaa
1191   aattcacact aaacaacaat gatcacagtt tcaatttcca aacattagca gacgcaacta attacgcttt
1261   aggtgtatag aaaaagaaaa tttctgatat taatgtatta gaagaaaaag aaatgcgtgc aatgttagtt
1331   gattactcat tgaatcaatt atccgaaaaa aatgtacgta aagcaacatc aaagaagat tagcaagca
1401   aagtttttga agcaatccta aacttacaaa acaacagtcg taaatataat gaagtacatc gtgcactcag
1471   tgggtgcaatt ggacaatata caactgtatc aaaaattaaa gatattgtga ttttaacaac agattcatta
1541   aaactttatc ttttagatag taagattgca aacacattcc agattgcagg cattgatttc acagatcacg
1611   tctattagttt cgacgactta ggtggcgtgt ttaaaagtaac aaaagaattt aagttacaaa accaagattc
1681   aattgacttt ttacgtgcgt atggagatta tcaatcacia ttaggagata caattccagt tgggtgtgta
1751   tttacttatg atgtatctaa acttaagagag tttactggca acgttgaaga aattaaacca aaatcagatt
1821   tatatgcggt tattttggat attaatctaa ttaaatataa acgttacaca aaaggtagt taaaaccacc
1891   attccataac cctgaatttg atgaagttac acactggatt cattactatt catttaagc cattagtcca
1961   tctttaaata aaattttaat tactgaccaa gatgtaaatc caaaaccaga ggaagaatta caagaataaa
2031   aggagcgtaa aatatgaaca acgataaaag aggtttaaac gttgagttat caaaggaat cagcaaaaga
2101   gttgttgaaac atcgcaacag attttaaact cttatgttta atcgttattt ggaattttta ccgctactaa
2171   tcaactatag caatcgatga acggttggta tagattttat tcagttagaa tcagctttta gacaaaacat
2241   taatgtagtt gttgggtgaag ctagaataaa gcaaatatag attcttgggt atgtaataaa cacttacttt
2311   aatcaagcac caaatttttc atcaaaactt aatttccaat ttcaaaaacg attaaactaa gaagatatat
2381   attttatgt acctgactat ttaatactg atgattgtct acaaatccat aagctatatg ataactgtat
2451   gagtggtaac ttgtgtgca tgcaaaataa accaattcaa tataatagtg atatagaat tatagaacat
2521   tatactgatg aattagcaga agttgtctta tctcgctttt ctttaactat gcaagcaaaa tttagcaaga
2591   tatttaaatc agaaattaat gacgagtcac tcaatcaact tgtgtccgaa atataatac gtcgaccatt
2661   tgttaaaatg tcacctatgt ttaatgcaga tgacgatatc attgatttaa caagtaatag cgtaatccca
2731   gcattaaactg aaatgaacg ggaatatcaa acaaaaatta gtgaattaa taactattta ggcattaaat
2801   caatagccgt tgataaagaa agcgtgtgtt cagacgaaga ggcaaaaagt aatcgtggat ttaccacatc
2871   aaacagtaat atctatttaa aaggtcgtga accaattacg tttttatcaa agcgttatgg tttagatatt
2941   aaacgtgatt acgatgatga aacaacgtct aaaatatcaa tggtagacac actttttaa gatgaaagca
3011   gtgatataaa tggctagata cacaatgact ttatacagatt tcattaaatc agaattgatt aaaaaaggtt
3081   tcaatgaatt tgtaaatgat aataaattaa cgttttatga tgatgaattt caattcatgc aaaaaatgct
3151   gaagttcgac aaagacgttt tagctatcgt taatgaaaaa gtatttaag gtttttcatt gaaagatgaa
3221   ttatcagatt tactttttta aaaaatcatt acgattcatt ttttagatag agaamtcaac agacaaacag
3291   ttgaagcatt tggcatgcaa gtgattactg tatgtattac acatgaggat tatttaaatg tggtttatc
3361   atcaagtga gttgaaaaat acttacaatc acaaggcttc acagaacaca atgaagatac aacaagtac
3431   actgatgaaa catcgaaatc aaatgctaca tctttagaca attcaactgg catgactgca aacagaacag
3501   tctatgtgtc attaccacaa agtgagggtta acattgatgt tgataatata acgttacgat tcgctgataa
3571   taatacagatt gataacggta aaactgtgaa taaatcgagt aacgaaagta atcaaaacgc aaaaacgtaat
3641   caaaatcana aaggtaatgc aaaaggtaca caattcacta agcagtattt aattgataat attgataaag
3711   cgtacgattt aagaaagaaa attttaaatg aatttgataa aaaaatgttt ttacaaattt ggtagaggtg
3781   gttaaataat ggcataataa gaaaacgatt ttaaatattt tgatgacatt cgtccatttt tagacgaat
3851   ttataaaaag agagaacggt atacacggtt ttacgatgat agagcagatt ataatactaa ttcaaatca
3921   tattatgatt atatttcaag attatcaaaa cttaattgaag tattagcacg tctgatttgg ccatatgaca
3991   atgaattaaa aaaacgtttc aaaaattggg acgacttaat gaaagcattt ccagagcaag cgaagactt
4061   atttagaggt tgggttaaacg acggtacgat tgacagtatt attcatgacg agtttaaaaa atatagcgca
4131   ggatcaacat cggcatttgc ttattttaa gttactgaaa tgaaacaaat gaatgacttt aaatcagaag
4201   ttaaagactt aattaaagat attgacggtt tctgttaatg gtttgaatta aatgagcttg aaccaaagtt
4271   tctgtatggg tttgggtgta ttgcgaacgc agttaaccaa tctattaata ttgataaaga aacaaatcac
4341   atgtactcta cacaatccga ttctcaaaa cctgaaggtt ttggataaaa taaattaaac cctagtgggt
4411   acttaatttc aagcatgcgt attgtacagg gtggtcatgg tacaacaatc ggattagaac gtcaatccaa
4481   tgggtgaaatg aaaatctggt tacatcacga tgggtgttga aaactgttac aagtcgcata taaagataat
4551   tatgtattag attagaaga ggttaaaagt ttaacagatt atacaccaca gtcactttta acaaaacaca
4621   catttacacc gtttaattgat gaagcaaatg acaaaactcat ttaagattc ggtgacgga caatacaggt
4691   tctgtcaaga gcagacgtaa aaaatccat tgataatga gaaaaagaaa tgacaattga taattcagaa

```

4761 aacaatgata atcgttggat gcaaggcatt gctgttgatg gtgatgatct atactgggta agtggtaaca
4831 gttcagttta ttccatggtt caaatcggta aatattcatt aacaacagggt caaaagattt atgattatcc
4901 atttaagtta tcatatcaag acggtattta tttcccacgt gataacttta aagagccctga gggattttgc
4971 atttatacaa atccaaaaac aaaacgtaaa tctgtattac ttgctatgac aaacggcgggt ggtggaaaaac
5041 gtttccataa tttatatggt tttctccaac ttggtgagta tgaacacttt gaagcattac gcgcaagagg
5111 ttccaaaaac tataaattta caaaagacga cggtcgtgca ttatctattc cagaccatat cgacgattta
5181 aatgacttaa cgaagcgtgg tttttattat attgacgggg gtactgcaga aaaacttaag aatagccaa
5251 tgaatggtag caagcgtata attgacgctg gttgtttcat taatgtatac cctacaacac aaacattagg
5321 tacgggtcaa gaattaacac gtttctcaac aggtcgtaaa atgggtaaaa tgggtcgtgg tatgacttta
5391 gacgtattta cgttaaaatg ggtatttgga ttatggacaa caatcaaaac tgacgcacca tatcaagaat
5461 atttggaaag aagtcataac aataactgga ttgcttatgt aacaacagct ggtgagttat acattacagg
5531 taaccaaatg gaattattta gagacgcgcc agaagaaatt aaaaaagtg gtgcatgggt acgtgtgtca
5601 agtggtaacg cagtcgggtga agtaagacaa acattagagg ctaatatatc ggaatatata gaattcttca
5671 gtaatgttaa tgcggaacaa aaacatcgtg ggtagcaaaa catcaaaat aggaagtata
5741 taaatgaant cacaacaaca agcaaaagaa tggatatata agcatgagggt ggcaggtgtt gactttgatg
5811 gtgcatatgg atttcaatgt attgacttat cagttgttta tgggtattac attactgacg gtaagttcgt
5881 catgtggggt aatgctaaag acgcatataa taatgacttt aaagggttag cgacgggtga taaaaatata
5951 ccgagcttta aacctcaatt aggggacggt gctgtatata caaatggaca atatggacat attcaattgt
6021 tgtaagttgg aaactctgat tattatacat gcttagaaca aaactgggtt ggcggcgggt ttgacgggtg
6091 ggaaaaagca accattagaa cacattatta tgacgggtga actcacttta ttgacctaa attttcagg
6161 agtaaataga aagcattaga aacatcaaaa gtaaatacat ttggaaaaatg gaaacgaana caatacggga
6231 catattatag aaatgaaat ggtacatttta catgtgggtt tttaacata ttgacacgtg tcgtagtcc
6301 aaaaattatca gaacctaatg gctattgggt ccaaccaaac ggttatcac catataacga agttttgtta
6371 tcatgattgt acgtatggat tgggtataac tggcaaggca caggttatta ttaccagtg cgcaaatgga
6441 atggaaaaac aggtcaatgt tacagtggtg gtattccttg ggggggtgtc tcataatggg tttttagcc
6511 ttttctttt aattttagtt gaaaagatac aataaagagg tgaatacaat tgaatacaat agcttaagaa
6581 gtttaagaca agttgaaaca attgaacggt tattggagga aaaaaatgag aaagtttaacg aatttttaag
6651 ttttctataa cacacggtt acagactatc aaaaacagat tcattttaat agtaataaag aacgtgtaga
6721 ttatttttta aatggtcgtc attttaaatc gttagactat tcaaaaacac cgtataaatt tatcgtgat
6791 agaattggaa tcaatgttga tatgcagtg catgacgcac aaggtattaa ctacatgacg tttttatcag
6861 attttgagga tagaagata tacygttttg taaaccaaat cgaatacgtg aatgacgttg tggtaaaat
6931 atattttgtc attgatacca ttatgacgta tacacaaggg aatgtattag agcaactctc aaacgtcaat
7001 attgaacgtc aaactttatc aaacgcgacg tataactata tgttaccat gttacgtaat aatgatgatg
7071 tgttaaaagt atcaaatata aactatggtt ataaccaaat gcaaccaat ttggaaaatt tagtattatt
7141 ccagtcgaagc gctgatttat caaagaaatt tgggtactaa aaagagccaa acttagatag gtaaaaaggt
7211 acgatttatg acaatatcac atccaccgtr aacttatag ttatggata tgggtacttt attaaactta
7281 tggataaaat gagtgccat ccatggatta cgcaaaactt tcaaaaaggt caaatgttac ctaaaagactt
7351 tattaatata aagacgttag aggacgttaa aaccagtgaa aaatttacag gattaaaaac attaaaaacg
7421 ggtggtaaat caaaagaatg ggtcctaaaa gatttatcat taagtttctc aaactctcaa gagatgagt
7491 tatctaaaaa agatgaattt aaacatatga tacgtaatga gatatgaca attgaatttt atgactggaa
7561 tggaaatacag atgttactcg acgctggtaa gatttcacaa aaactcgtgt ttaagttacg tacaataatc
7631 attattgggt atcataatga agttcgagta tatccagtag attataacag tgcgtaaaac gacagacca
7701 tactcgctaa aaataaagaa atattgattg atacgggttc attcttaaat acaaatatata catttaatag
7771 ttttgacaaa gtaccaatat taatcaataa tggatcttta ggacaatcac aacaagccaa ccgacaaaaa
7841 aatgcgaaaa gtcaactaat tacaaaactg tacaatatcg tattaaatgy tattaataggy taagcaccgct
7911 tttatgacgc tgtgagtgta gcaagtaatt taagtccaac tgctttatgt ggttaagttta atgaagata
7981 taattttctac aaacaacac aagctgaata taaagattta gccttacaac caccctctgt atgaagata
8051 gaaatgggga acgcatcca aacttcgaat agcattraacg gtttaacgag gtaaaataggt gtaacgtcac
8121 ctaaaagaaat tacattttta caaaaatatt atatgttgtt tgggtttgaa gtgaatgact ataatcatt
8191 tattgaacca attaacgta tgactgtttg caattattta aaatgtacag gtacgtatag tatcgtgac
8261 atcgacccca tgttaatgga acaattataa gcaatttttag aatctgggtg aagatttttg cataatgacg
8331 gttcaggtaa tccaatgta caaaatccat taaataacaa atttagagag ggggtataat atgaacgaag
8401 taaaattcag atttacagac tcagaagcgt ttccatggtt tatatacgtt ggggatttaa aattactcta
8471 cttttttatt gtattaatgt tcgttgatat tattacaggt atttcaaaag caattaaaaa taataactta
8541 tgggtcaaaa aatcaatgag aggtatttct aaaaaattat tgatattctg tattatcatt tttagcaaca
8611 tcattgacca gattttacaa ttaaaagggt gttactcat gattacaata ttttattata ttgcaaatga
8681 gggactttct attgtagaaa attgtgcaga aatggacgta ttagtaccag aacaaattaa agataaatta
8751 agagtcatta aaaaatgac tgaanaaggt gataacaaat aacgatcaag agaagataga caaatttacg
8821 cacttctata ttaactgata ctttgggtta acgatagacc agttagtccc taaagtaaaa ggtatggg
8891 gctttaaatt atggcttgggt ggttaagaaa gtaaaatcag acaagtatta aaagcagtaa aagagatagg
8961 tgtttcacct actctttttg ccgtatatga aaaaaatgag ggttttagtt ctggacttgg ttggttaaac
9031 catacgtctg cacgtgggtg ttatttaaca gatgctaaat tcatagcaag aaagttagta caaatggg
9101 aacaagctgg acaaccgtct tggttatgac caggttaacat cgtccacttt gtaccacaag acgtacaaa
9171 aaaaaggtaat gcagattttg caaaaaatat gaaagcaggt acaattggac gtgcatatat tccattaaac
9241 gcagctgcta cttggggcgg atattatcct ttatgtttga aagcatcata taacaaagta caaaactatg
9311 gtaatccatt tttagacggt gcaataacta ttctagcttg ggttggtaaa tttagcggta aaggtggatc
9381 acctagtgat tctgtgaca gtggtagtag tgggtgacag aatgggacgt tcatagttat ggtagtgata
9451 gccatgcacg aattattaaa aaaaatacaa gacgcattac aatgggacgt tatcatatta aaatgacgt
9521 aatttttttag taatgattat ttacatttag aaaaaacatt caagtagata tggggagtag tagttctaac
9591 tgggttactt gattcattaa aaaaactgat tgatagcgtt caagtagata ggggagtag aaaaagtggt
9661 cctactgatg atgacggaga ccataaaacca attagtggta aatcagtcac aaagcaattg gtgtacctt
9731 gtgtgatttg tggtaactgg acatattgac agttaccaga aaaaataaaa gtaatgcagg acaatgtaca
9801 attcaaaaaa gaatacttat acaaacagg taacatat ttcatagctt cctcaaacgg gtaatgcagg
9871 gaattaacat gggcgtatat gtcacaacta catggtaaaa gacaacctac cgacgcaggt caaataacaa
9941 acggtcagcg tgtatggtag gtctataaaa agttagggtg aaaaaacaca cataatccaa cagttagtta
10011 tgggttctct agtaaacacc catacttaca agcaactgca tatgggtatt

10081 gcagtttttg aagatgggtc gtttttagtt gcaaaactata atgtaccacc atatgttgca ccatcacgtg
 10151 tgggtattgta tacactcatt aatggcggtac caaataatgc tgggtgataat actgtattct ttagtgggtat
 10221 tgcttaatta actatgctat aatgaacaca tgctagtaat gctagtataa aaaaacacaa acataatcaa
 10291 ttttcgtaca cttttttcat gttatctcaa aaagaaaagg agactgttat tttaacagtt gctttttttt
 10361 atttcatcat gttcacgttt taatatatgc aaatcagatt tgttatgtac tgaacgttca actggaaata
 10431 agtcgttaag tgaanaatgaa cggatgtcac tttcaatata aagaatatca tcaaatgtac tatgggtcgaa
 10501 attttctcta ggcgtcttta atataaaatc acgtttcata ttaagttcat cagtaaaata ttcacatcat
 10571 acattaccac atacaatttc agtttttagac ggataatcg atattgtacc ttgtcatta tagatacttt
 10641 tactgttttc aataatggca cgttcaaga attgttcacg tacaagggtt tcaaatcga cgtttgtatc
 10711 aaaggcggtt ttcgggtatc cagcagaagc aattttaatc tttccattca cttcatatgc atattttcta
 10781 tgattcagta caaacatctt atctatctgt tegttttcaa tatccattt acctaaggct atcgggtcga
 10851 ataaactggg gttcaataag ggtttaacaa cggatttcat atacaaacta tcagtatcgc aataaataaa
 10921 attgtcgta atttcacttt cgtttaagta ttggaaggga accaataagt tatacaatga acgtgatgtg
 10991 acaaatgtag agaataatat attacgttca gtgtttttgt aaaccttaat gatattgtat agttcattgt
 11061 tatcatctaa acggaataag ttaaaatgtg aacgttaatgc aggtatgcca tataatccat ttaaaacgac
 11131 tttagataac ataactcctt catttgagta tgggtgtctg ttgatcatcat cagtaagtgt atagtgtaa
 11201 ggtgatgtca tattgatttt gttttttaac ttaccttggt ttttaataaa atagttttga aaaaataat
 11271 cacgtgcatg aaagtattca cattcatata taacaaacga attaaacagt atatgcatgc aaacataacc
 11341 cgttaaatgt tgaatcattc ttaatgtatt tgtattgata ttaacgtaat cattatcatt attatagtat
 11411 tttacaatca tttagcgtta tacacgtgat ttaattttaa ttaataaatc atcgttaaat acatctttat
 11481 caatcttata taatgaaaaa taattgtcat catctaaaaa agcagggatt aacgttgggt ctgaatagtg
 11551 ttctgtaaaag tataaccatg ttggaatttt ttcatgata atcacataag gataactcga attgatgtca
 11621 atagaaaaac aagggtcctc aattagtttg tttatgtatt tgggtgtata catattttaa cccaccgat
 11691 agaattgatt aatcatagta taaaatttca tatcatggaa atgataatgt gtataagata ttttaatatc
 11761 ttgatatttg ttgagtaact gaaaacgtgt catttcatta ttcaagtaag attccataat attcaatgaa
 11831 aatgttaatt tgttatagtc aaaaatttga aatatatcac tataatgaat atggcacata cctaataata
 11901 tcacgtcatt atgaatgtat gtaagtgtgt cagggtgtgag ttttgcaaaa catttcacag catagtcata
 11971 ggcttcacta tcatcatatc cattatcttt atcaaaaatc gtataattaa aatctgtttt aagttgtgat
 12041 tctgttaaat aaccaccatc aagtaatttc ttacctaattg ttgcaattga tgtattgggt ttcataaagt
 12111 tatcaataat attaaattta aaaccattta aaacatttgt taactctaaa ttgattgaag atttaacacg
 12181 tttttctaaa attacatttt gatttttggc taataatagta gcttctttca tttttaatgt ttgttcat
 12251 tcttctcgag atttttaata tatattttcg cgtgtaatat tatcaaaaata acgcatgggt tctttaagta
 12321 aaaaatgatt atcgtattta ttacagttat gtgcaatcat gataatctct gtttttgatt ttgtgattgt
 12391 atcactgctt ttacacatgc tataaaatgc gtcatataaa gattcgaaac ttggaataac ttcaacatca
 12461 atttcataac cattaaacca accaattgct acagaataag taacgttttt atatttgggt ggttttttct
 12531 gtcctgttaac tttattgtac gctaatgttt ctatatccca gtataaaatc attcgaactt catgtttatg
 12601 atactgcatg cattctagta atcccataat cttacacacc ttttataagc catattgttt cattagatc
 12671 tttttcgat tctctatata gttatctctg tatatttttt cttttcttcc aaactcactc atatttttct
 12741 tcatcttcat ttttatatga aattttataa tttttattcat atctaaatat aaatatctat cattatcaac
 12811 cactgaattt ttagagtaag cattgtcaaa atgtaaattg cttggattgt agtaataacg ttccattgtt
 12881 tctttatata acatatcctc acgttaaatag gtaacatgat tgtctatctc cctaatttta gtacaaatt
 12951 catattgttt tgtatatggt acaacgataa tatttgcatt aaaaagtagt acattataca tgactttat
 13021 atatttaca tcaattttga tatagaagaa atcaccgttt tgattgatgt gatttcttaa attatcatcc
 13091 gccaaattat attcgtttaa ttcaaatctc ccagttgtca tagcgtgtgc atttgaatta aacgcacgtg
 13161 tgttactgtt ttcatctcag taatcgtttc gtgcatttcc taataaaatg tttttgtaaa gttttgatgt
 13231 attcatttta tgcctttgta ataaattgta tatattttaa ttggataata taggacttga aaagttagt
 13301 gcattaccta gtaaaaacat tttagggaat ccaataataat caacgttacc atggttacgg tcgattgat
 13371 catatttgt ttttaactta tccactcact caattaaata atcatcttca agtgctaaaa actcatcaca
 13441 tatataata ggatagtggt ttaaaaagtt agaattgat tttaaatcag tggcactatt caaatctgta
 13511 atcacacca tttctttatc ttgatagata atagctaaat agtccctagc acttctgaac gtgacacgtt
 13581 ttgattttaa tagtggattt tctatctatg tttcttcaat aaatcacggt taagcgtcac gtaactgata
 13651 attgacgtgt aataaagtaa attttatctc aagtttaata gctaaataaa taataaaatga aacatagttg
 13721 aacgtttttc catcagaacg gtttgaataa gatataat aatctatctc atcattcata agttcatcaa
 13791 ctaattctat ttgattatc ttatctggga tttttttctt gacatgattg acagcatttt gataatctct
 13861 taccatgtct aaacgttttt gttttaccat gtttttgctc cttgtaatag ttcatgatgt cgtttacagt
 13931 gttaaattta ttctgcaaat gttgcataat ataaaagtt acacctcaca tcttcatcat caatatttgt
 14001 cactggctta tctgatttac caattctttt atataaagta tctgattctt taatatattt atacattgaa
 14071 gaattattat ttttagcttg taattatata aaacgttatt tatgcttttt agcgttttta ttattagaa
 14141 catcattacg gtttatatt tcaagaatat aatttaattt tttatgtctt gaacctctta ccaatgatc
 14211 agcatttaca tatgatcgt tttctttctt aggaataatg ggcagatgtg caaatgtttt ccatgtgtca
 14281 atgtacgctt cttgttaatc tttatcatca aatttaaaat taactattact aaatcattt aaaaataaat
 14351 cttttttctg ctcttttcta gcttctcttt cttttttcca tctatccatt tcagacgtat gtttaaccac
 14421 tgttatcaac ctccatataa agcataaata accattaaaa agataatata gaataatac aatgtagtga
 14491 ataaaacacc aaatgacacg cgttatatga gtgtcataag tatgataagt gtaattaaaa atgctaaaag
 14561 gaaaacaatg gctatgttta ataggttatt catgggtcaat cactttccca tctatcgata tgactttgtt
 14631 ttgataaata atcatttaatt cgtcttcaag aggtttatca aaatttgata atacgtcgtc aatgttaacg
 14701 ttttaaaaaa tttctcttat taattcatta ctttaataat ttttataata aaatacaagt atattaaaa
 14771 catgtttttt aatatcaatg tcgatatcta acgttaataa ctttttttca atttcaaat catcatattg
 14841 tttgtcaaac tcaatatata catcaccat atttattttt actatacatt ttttattaga tgaagttaat
 14911 ttttcaaat tatcattata ataatctcta tttgttaaaa ggtataaat taaattattt aatctaaaag
 14981 tagttttatt tttcattttt atatctctt atgtattctt atgatatcag cgtatttttt agtgaatagg
 15051 tatattctat aatatgaata tacaacttta gcgtcatata aatcttcaaa cattgagatt tgatgtggaa
 15121 aatgtctctt aatctcatcg caatataata ataccgtttt gtatttacgt tccattttaa cactcataa
 15191 aaaaatgggg ataatgtatcc cctatgaat tgaattaaaa tgatacttga ccaaatgta ttgagtaacc
 15261 tttttgacct tttttgtttt catattcata aattgtgaat tgaacttctc cagcattgat aatgtcaaca
 15331 acgtctcat ctgctctcat tttttaatt aattctgtta agtgggtcgg taagtttacg ttatagcat

269

```
15401 cagtgcacgat aacacccctgt tcaccgaatt ttgattcttt gtttgtgaat aatgctctaa cgatatactc
15471 ttttttcata ccgtattttt ctactaatte tgatagtttg ataaattctc tttctttttc ctcaaatcca
15541 aatcttcgcta atgtgttttg gctgtttgat aaaatatctt ttacgtttgc cattttattt ctcctcttat
15611 ttaaatattt tgctttctgc aattgctgatt tgtagttaaat cattgttaata aacttgaatt gttttcgttg
15681 tgcgtgtagt ggacaatagt ttacatgtgt crggtataaa ttcttttgct tgtgttttgg ttaaatgata
15751 ctctggaagt ggtaaaaatt cctcaatgta ttcatatca tcatcctaagt aatgaagtat ataacccttg
15821 acacgtaagg taacaatgtc gtcaactttc attattatat cactcctttc taaaaaacgt aaacgttata
15891 cgtttcataa aatcctttat gcataattcca ttgttctatt gggtcatac cagcaatata agacaatatt
15961 gattcttggt tagtttcgtt gtttagttca tcatttaaga attgaacaac agaactatta tagtttaata
16031 atagtgttg gcaagccgat aataagttaa ttgcattgtc aaatgtataa gctggattcc attgaatcag
16101 tttattgaat agttgcaaca ttccagtata ggcttgctct tttcttctg gtgcattatc aacattaacc
16171 attattatca cttcctaata aagttgaaat tacgcgtaaa acagaattat gatttaaatc ttcattttca
16241 tcaatgtcaa catcataaaa tgaaatttca tttctgttc tatcaataaa cgctatacat aaacttccat
16311 tcttaaaacg aaaaacatgc ttcaactcaa tgttttttgt ttcattttcc atttttgta ctccttggtt
16381 tgattacata cttagtatag caaacgttta aaagttttgt caatagtttt tctraaaaaa gtttaataaa
16451 ttttaaaact actatttaat agaagaaata agatttttaag ttcaaatcat aattttgaat aaaagtcatt
16521 agatacataa attttgtatt tgatgaatat gtaatagggt agataagttg gtttaagttg tgcacagtat
16591 ttttaagttt agtaaagaaa tgataagtaa atttataagt ttgtatttgc ataactgttt atttcaaac
16661 ggtgggggt
```

Table 17

Phage 44AHJD ORFs list

nb	Name	Frame	Position	Size (a.a.)	Key words
1	44AHJDORF001	-1	10342..12627	761	DNA polymerase;
2	44AHJDORF002	3	3789..5732	647	Teichoic acid; Staph;
3	44AHJDORF003	2	6626..8389	587	Tail;
4	44AHJDORF004	1	8764..10227	487	Serine protease motif;
5	44AHJDORF005	-1	12643..13890	415	
6	44AHJDORF006	2	803..2029	408	
7	44AHJDORF007	1	2044..3027	327	Upper collar;
8	44AHJDORF008	2	3020..3775	251	Lower collar;
9	44AHJDORF009	2	5744..6496	250	Amidase; Staph;
10	44AHJDORF010	-2	13938..14420	160	
11	44AHJDORF012	3	8391..8813	140	Holin;
12	44AHJDORF013	-2	14586..14996	136	
13	44AHJDORF113	1	199..600	133	
14	44AHJDORF011	-2	15225..15593	122	
15	44AHJDORF114	-2	15870..16172	100	
16	44AHJDORF014	3	6243..6521	92	
17	44AHJDORF015	1	15403..15645	80	
18	44AHJDORF016	-1	15616..15852	78	
19	44AHJDORF017	-2	10536..10757	73	
20	44AHJDORF018	-1	886..1098	70	
21	44AHJDORF019	-2	9630..9836	68	
22	44AHJDORF121	-1	16165..16362	65	
23	44AHJDORF020	2	13865..14053	62	
24	44AHJDORF123	2	614..796	60	
25	44AHJDORF021	-2	5634..5816	60	
26	44AHJDORF023	-2	6315..6494	59	
27	44AHJDORF024	1	14275..14451	58	
28	44AHJDORF025	-3	14999..15175	58	
29	44AHJDORF026	-3	14426..14593	55	
30	44AHJDORF027	1	12916..13080	54	
31	44AHJDORF029	-1	15019..15183	54	
32	44AHJDORF028	-3	9071..9235	54	
33	44AHJDORF030	3	14487..14648	53	
34	44AHJDORF031	2	11039..11191	50	
35	44AHJDORF135	3	693..842	49	
36	44AHJDORF033	-1	3648..3795	49	
37	44AHJDORF032	-2	9306..9455	49	
38	44AHJDORF034	-3	14000..14146	48	
39	44AHJDORF035	-3	13811..13957	48	
40	44AHJDORF036	-3	10019..10165	48	
41	44AHJDORF022	-3	8488..8611	47	
42	44AHJDORF037	1	14788..14931	47	
43	44AHJDORF038	-2	3528..3671	47	
44	44AHJDORF039	3	1743..1883	46	
45	44AHJDORF040	2	9740..9877	45	
46	44AHJDORF041	2	15838..15973	45	
47	44AHJDORF042	-1	5014..5151	45	
48	44AHJDORF043	-1	4402..4539	45	
49	44AHJDORF044	-2	12783..12917	44	
50	44AHJDORF149	-2	639..770	43	
51	44AHJDORF046	1	4891..5019	42	
52	44AHJDORF047	1	11911..12039	42	
53	44AHJDORF045	2	10655..10783	42	
54	44AHJDORF048	-3	15212..15340	42	
55	44AHJDORF049	3	5784..5909	41	
56	44AHJDORF050	3	13158..13283	41	
57	44AHJDORF051	-2	10944..11066	40	
58	44AHJDORF052	-3	14216..14338	40	
59	44AHJDORF053	3	3348..3467	39	
60	44AHJDORF054	3	7551..7670	39	
61	44AHJDORF055	3	15705..15821	38	
62	44AHJDORF056	1	5512..5625	37	
63	44AHJDORF057	2	10121..10231	36	
64	44AHJDORF058	3	10767..10877	36	

271

65	44AHJDORF164	-1	592..702	36	
66	44AHJDORF059	-2	8250..8360	36	
67	44AHJDORF060	-2	6147..6257	36	
68	44AHJDORF061	2	15551..15658	35	
69	44AHJDORF062	1	4285..4389	34	
70	44AHJDORF063	-3	9383..9487	34	
71	44AHJDORF065	1	5029..5130	33	
72	44AHJDORF064	2	2609..2710	33	
73	44AHJDORF066	-2	10380..10481	33	

Table 18

Predicted amino acid sequences

44AHJDORF001

```

12627 atgggattactagaatgcacgaatcacaacatgaacgtcgaaatgattttatactgggatagaaacattagcgatcaat
1 M G L L E C M Q Y H K H E R R M I L Y W D I E T L A Y N
12543 aaagttaacggagcaaaaaaaccaacaaatataaaaacgttacttattctgtagcaattgggtgggttaattgggtatgaaatt
29 K V N G R K K P T K Y K N V T Y S V A I G W F N G Y S I
12459 gatgtgaagtatttccgaggttccgaatcttttatgacgcattttatatacgtatgtgaaagacgtgatacaatcacaaatca
57 D V E V P P S P E S F Y D A F Y T Y V K R R D T I T K S
12375 aaaacagatattatcatgattgcacataactgtaataaatacgataatcatttttacttaagacaccatgcgttattttgat
85 K T D I I M I A H N C N K Y D N H F L L K D T M R Y F D
12291 aatattacacgcgaaaaatataatatttaaaactcgcagaagaaaatgaacacacattaaaaatgaaagaggctactatttttagcc
113 N I T R E N I Y L K S A E E N E H T L K M K S A T I L A
12207 aaaaatcaaaatgtaattttgaaaaaacgtgttaaatcttcaatcaatttagatttaacaatgtttttaaattgggttttaattt
141 K N Q N V I L E K R V K S S I N L D L T M F L N G P K F
12123 aatattatgataactttatgaaaaaccaatcacatcaattgcaacattaggtaagaattacttgatgggtgggtatttaacagaa
169 N I I D N F M K T N T S I A T L G K K L L D G G Y L T E
12039 tcacaacttaaaacagatttttaattatcagatttttgataaagataargatgaatgatagtgaaagcctatgactatgctgtg
197 S Q L K T D F N Y T I F D K D N D M N D S E A Y D Y A V
11955 aaatgttttgcacaaactcacacctgaacaacttacatcattcacaatgacgtgattatattaggtatgtgcatattcattat
225 K C F A K L T P E Q L T Y I H N D V I I L G M C H I H Y
11871 agtgataattttccaaattttgactatacaaaatcaacattttcattgaattatattggaatcttacttgaaatgaatgaca
253 S D I F P N F D Y N K L T F S L N I M E S Y L N N E M T
11787 cgttttccagtactcaaccaatatacaagatattaaaatattctatcacattatcatttccatgatatgaatttttatgactat
281 R F Q L L N Q Y Q D I K I S Y T H Y H F H D M N F Y D Y
11703 attaaatcattctatcgtgggtgtttaaataatgtataacaccaaatacataaaacaaactaattgatgagcctgtttttctatt
309 I K S F Y R G G L N M Y N T K Y I N K L I D E P C F S I
11619 gacatcaattcagagttatccttatgtgatgtatcatgaanaaaatccaacatgggttatacttttgcgaacactattcagaacca
337 D I N S S Y P Y V M Y H E K I P T W L Y F Y E H Y S E P
11535 acgttaatccctacttttttagatgatgacaattatttttattatataagattgataaagatgtatttaacgatgatttatta
365 T L I P T F L D D D N Y F S L Y K I D K D V F N D D L L
11451 attaaaattaaatcagtgattacgtcaaatgattgttaaaataactataaataatgataatgattacgttaatatcaatacaaat
393 I K I K S R V L R Q M I V K Y Y N N D N D Y V N I N T N
11367 acattaaagaatgattcaagacattacgggtattgattgcacatatacgtgttaattcgtttgttatatatgaatgtgaatac
421 T L R M I Q D I T G I D C M H I R V N S P V I Y E C E Y
11283 tttcatgcacgtgataattatttttcaaaactattttatataaacacaaaggtaagttaaaaaacaaaatcaatgatgacatcacct
449 F H A R D I I F Q N Y F I K T Q G K L K N K I N M T S P
11199 taacgatcacattactgatgatatacaacgaacacccataactcaaatgaggagggttatgttatcctaaagtcgttttaaatgga
477 Y D Y H I T D D I N E H P Y S N E E V M L S K V V L N G
11115 ttatattggcatacctgcattacgttcacattttaaacttattccgttttagatgataacaatgaactatacaatattcattacgggt
505 L Y G I P A L R S H F N L F R L D D N N E L Y N I N G
11031 tacaaaaaacactgaacgtaatatattattctctacatttgcacatcacgttcattgtataacttattgggttcccttccaatc
533 Y K N T E R N I L F S T F V T S R S L Y N L L V P F Q Y
10947 ttaacggaaagtgaattgacgacaatttttattttattgcgatactgatagtttgtatatgaaatccggttggttaaacctttattg
561 L T E S E I D D N F I Y C D T D S L Y M K S V V K P L L
10863 aaccccggtttattcgacccgatagccttaggttaattgggatatgaaaacgaacagatagataagatgtttgtactgaatcat
589 N P S L F D P I A L G K W D I E N E Q I D K M F V L N H
10779 aagaaatagcatatgaagtgaatggaagattaaaattgcttctgctgtataccgaaaaacgctttgatcaagcgtcgat
617 K K Y A Y E V N G K I K I A S A G I P K N A F D T S V I
10695 tttgaaacctttgtacgtgaacaattctttgacggtgccattattgaaaacaataaagatctataatgagcaaggtacaata
645 F E T P V R E Q P F D G A I I E N N K S I Y N E Q G T I
10611 tcgatataatccgtctaaaactgaaattgtatgtggtgaatgtatgatgaatattttactgatgaacttaatatgaaacgtgaa
673 S I Y P S K T E I V C G N V Y D E Y P T D E L N M K R E
10527 tttatattaaaagacgctagagaaaatttcgaccatagtcattttgatgatattctttatattgaaagtgcacatcggttccattt
701 F I L K D A R E N F D H S Q F D D I L Y I E S D I G S F
10443 tcacttaacgacttatttccagttgaacgttcagttacatacaaaatctgatttgcataattaaaacgtgaaacatgatgaaata
729 S L N D L F P V E R S V H N K S D L H I L K R E H D E I
10359 aaaaaaggcaactgttaa 10342
757 K K G N C *

```

44AHJDORF002

```

3789 atggcatataatgaaaacgatttttaaatattttgatgacattcgtccatttttagacgaaatttataaaacgagagaacgtttat
1 M A Y N E N D F K Y F D D I R P F L D E I Y K T R E R Y
3873 acaccgtttttacgatgatagagcagattataatactaatccaaatcatattatgattatatttcaagattatcaaaactaatt
29 T P F Y D D R A D Y N T N S K S Y Y D Y I S R L S K D I
3957 gaagtattagcagcgtcgattttgggactatgacaatgaattaaaaaacggtttcaaaaattgggacgacttaataaagcattt
57 E V L A R R I W D Y D N E L K K R P K N W D D L M K A F
4041 ccagagcaagcgaaagacttatttagaggttggttaaacgacggttacgattgacagattatttcatgacggttttaaaaaatat
85 P E Q A K D L F R G W L N D G T I D S I I H D E F K K Y
4125 agcgcaggattaaacatcggttatttcttttaagttaactgaaatgaaacaaatgaatgacttttaaatcagaagtttaagac
113 S A G L T S A F A L F K V T E M K Q M N D F K S E V K D
4209 ttaattaaagatattgacggtttcgttaattgggttgaattaaatgagcttgaaccaaagtgtgtggtgggttgggtgatt

```


141 L I K D I D R F V N G F E L N E L E P K F V M G P G G I

4293 cgcaacgcagtgtaacaaatctattaattgataaagaacaaatcacatgtactctacacaatccgattctcaaaaacctgaa
169 R N A V N Q S I N I D K E T N H M Y S T Q S D S Q K P E
4377 gggttttgggataaataaattacacctagtggtgacttaatttcaagcatgcgtattgtacaggggtggtcattggtacacaatc
197 G P W I N K L T P S G D L I S S M R I V Q G G H G T T I
4461 ggattagaacgtcaatccaatgggtaaatgaaatctgggtacatcacgatgggtgtgcaaaactgttacaagtgcgcatataaa
225 G L E R Q S N G E M K I W L H H D G V A K L L Q V A Y K
4545 gataattatgtattagatttagaagaggctaaaggcttaacagattatcacaccacagtcacttttaacacaaacacacattaca
253 D N Y V L D L E E A K G L T D Y T P Q S L L N K H T F T
4629 ccgttaattgatgaagcaaatgacaaactcattttaagattcggtgacggaacaatacagggttcgttcaagagcagacgtaaaa
281 P L I D E A N D K L I L R F G D G T I Q V R S R A D V K
4713 aatcacattgataatgtagaaaagaaatgacaattgataattcagaaaaaatgataatcggtggatgcaaggcattgctgctt
309 N H I D N V E K E M T I D N S E N N D N R W M Q G I A V
4797 gatgggtgatttatactgggttaagtggttaacagttcagttacatgttcaaatcggttaaatattcattacaacaggt
337 D G D D L Y W L S G N S S V N S H V Q I G K Y S L T T G
4881 caaaagatttatgattatccatttaagttatcatatcaagacgggtattatctccacgtgataactttaagagcctgagggt
365 Q K I Y D Y P F K L S Y Q D G I N F P R D N F K E P E G
4965 atttgcatttatacaaatccaaaaacaaaacgttaaatcggttacttactgtatgacaaaacggcggtggtggaaaacgtttccat
393 I C I Y T N P K T K R K S L L L A M T N G G G K R F H
5049 aatttatattgggtttcttccaaactgggtgagtargaacactttgaagcattacgcgcaagggttcacaaaactataaaataaca
421 N L Y G F F Q L G E Y E H P E A L R A R G S Q N Y K L T
5133 aaagacgacggctgcgttcattatctattccagaccatctcgacgatttaaatgacttaacgcaagcgtggtttttattatattgac
449 K D D G R A L S I P D H I D D L N D L T Q A G P Y Y I D
5217 ggggttactgcagaaaaacttaagaatatgccaatgaattggttagcaagcgtataattgacgctggttcttcaattatgatac
477 G G T A E K L K N M P M N G S K R I I D A G C F I N Y Y
5301 cctacaacacaaacattaggtacgggttcaagaattaacacgtttctcaacaggtcgtaaaatgggttaaaatgggtgcgtggtatg
505 P T T Q T L G T V Q E L T R F S T G R K M V K M V R G M
5385 ttttagctattttacgttaaaatgggtattatggattgggacaaacaaactgacgcacacatatacaagaattatttggaa
533 T L D V F T L K N D Y G L W T T I K T D A P Y Q E Y L E
5469 gcaagtcatacaataactggattgcttatgttaacaacagctgggtgagtattacattacaggttaaccaaatggaattattttaga
561 A S Q Y N N W I A Y V T T A G E Y Y I T G N Q M E L F R
5553 gacgcgcagaaagaataaaaaagtggtgctggttactgtgtaagtggttaacgcagtcggtgaagtaagacaaacatta
589 D A P E E I K K V G A W L R V S S G N A V G E V R Q T L
5637 gaggctaatatcggaataataagaattcttcagtaattgttaatgcggaacaaacacatcgtaattatgggtggtgacaaa
617 E A N I S E Y K E F F S N V N A E T K H R E Y G W V A K
5721 catcaaaaatag 5732
645 H Q K *
44AHJDORP003

6626 atgagaaagttaacgaatttttaagtttttctataacacacgcgtttacagactatcaaaacacgattcattttaatagtaataaa
1 M R K L T N F K F F Y N T P F T D Y Q N T I H F N S N K
6710 gaacgtgatgattatttttaaatgggtcgtcattttaaatcggttagactattcaaaacacacgtataattttatcagtgataga
29 E R D D Y F L N G R H F K S L D Y S K Q P Y N F I R D R
6794 atggaatcaatgttatgacgtggcatgacgcacaggtatttaactacatgacgtttttatcagattttgaggtatagaaga
57 M E I N V D M Q W H D A Q G I N Y M T F L S D F E D R R
6878 cattacgcttttttaaaccaaatcgaatcgtgaatgacgttgggttaaaatataatttgcattgatacattatgacgtat
58 Y Y A F V N Q I E Y V N D V V V K I Y F V I D T I M T Y
6962 acacaagggaatgtattagagcaactctcaaacgtcaaatattgaacgtcaacattttatcaaaacgcacgtataactatattgta
113 T Q L N V L E Q L S N V N I E R Q H L S K R T Y N M L
7046 ccaatttactgtaataatgtgtttaaagattcaataaaaaactatgctttatacccaaatgcaacaataatttgggaat
141 P M L R N N D D V L K V S N K N Y V Y N Q M Q Q Y L E N
7130 ttagtattattccagtcgaagcgtgatttatcaagaaattttggtactaaaaagagccaaacttagatagctcaaaagggtacg
169 L V L F Q S S A D L S K K P G T K K E P N L D T S K G T
7214 atttatgacaatatcacatcacagtcacacttatcgtttatggaatatggtgactttattactttatggataaaatgagtgcc
197 I Y D N I T S P V N L Y V M E Y G D F I N F M D K M S A
7298 tatccatggattacgcacaaactttcaaaagggttcaaatgttacctaagacattttattaatacaaaagacttagaggcgttaaa
225 Y P W I T Q N F Q K V Q M L P K D F I N T K D V K
7382 accagtgaaaaaattacaggattaaaaacattaaaaacagggtggttaaatcaaaagaatggagtcataaagatttatcattagt
253 T S E K I T G L K T L K Q G G K S K E W S L K D L S L S
7466 ttctcaaatcttcaagagatgatgttatctaaaaaagatgaatttaacatagatagcgaatgagtatatgacattatgaaattt
281 F S N L Q E M M L S K K D E F K H M I R N E Y M T I E F
7550 tatgactggaatggaatacagatgttactcgacgctggtaagatttcacaaaaactgggtgtaagtacgtacaaaactcaatt
309 Y D W N G N T M L L D A G K I S Q K T G V K L R T K S I
7634 attgggttatcataatgaagttcgagttatccagtgattataacagtgctgaaaacgcagacacaaactcgttaaaaaataaa
337 I G Y H N E V R V Y P V D Y N S A E N D R P I L A K N K
7718 gaaatattgattgatacgggttcattcttaatacaaaataaacatttaattagttttgcacagttacaaatattatcaataat
365 E I L I D T G S F L N T N I T F N S F A Q V P I L I N N
7802 ggtatcttaggacaatcacaaacagccacacgacaaaataatgcagaaagtcaattaattacaacacgtattgataatgattat
393 G I L G Q S Q Q A N R Q K N A E S Q L I T N R I D N V L
7886 aatggtagcaccgcaaatcacgcttttatgacgctgaggtgtagcaagtaatttaagtcacactgcttatttggtaagttt
421 N G S D P K S R F Y D A V S V A S N L S P T A L P G K F
7970 aatgaagaataataattctacaacaaacaaacagctgaataaagatttagccttacaaccacctctgtaactgaatcagaa
449 N E E Y N F Y K Q Q A E Y K D L A L Q P P S V T E S E
8054 atgggcaacgcattccaatttcgaatagcattacgggtttaaagcagtgaaaatttagtgacgtcaccttaagaagaattacatt
477 M G N A P Q I A N S I N G L T M K I S V P S P K E I T F
8138 ttacaaaaattattatggttgggttttgaagtgaaatgactataattcatttattgaaccaattacacgatgactggttgc

505 L Q K Y Y M L F G F E V N D Y N S F I E P I N S M T V C
8222 aattattttaaaatgtacaggtacgtataactatcgtgacatcgaccccatgttaaatggaacaattaaaagcaatttttagaatct
533 N Y L K C T G T Y T I R D I D P M L M E Q L K A I L E S
8306 ggtgtaagattttggcataatgacggttcaggtaacccaatgttacaaaatccattaaataacaaatttagagaggggtataa
8389
561 G V R F W H N D G S G N P M L Q N P L N N K F R E G V *
44AHJDORF004
8764 atgatactgaaagagtgataaacaatgaacgatcaagagaagatagataaatttacgcattcctatatattaatgatgattttggt
1 M I L K R V I T M N D Q E K I D K F T H S Y I N D D F G
8848 ttaacgatagaccagttagtccttaaaagtaaaaggatagggcgcttaagtgtatggcttgggtggaatgaaagtaaaatcaga
29 L T I D Q L V P K V K G Y G R F N V W L G G N E S K I R
8932 caagtattaaaagcagtaaaagagataggtgttcctactctcttttgcggtatgtgaaaaaatgaggggttttagttctgga
57 Q V L K A V K E I G V S P T L F A V Y E K N E G P S S G
9016 cttggttgggttaaacaccatcgtctgcacgtggtgattatttaacagatgctaaattcatagcaagaagtttagatcacatca
85 L G W L N H T S A R G D Y L T D A K F I A R K L V S Q S
9100 aaacaagctggacaaccgtcttgggtatgacgcaggttaacatcgctccactttgtaccacaagcgtacaaaagaaaggttaatgca
113 K Q A G Q P S W Y D A G N I V H P V P Q D V Q R K G N A
9184 gattttgcaaaaaatgatgaaagcaggtacaattggcagtgcatatttccattaacagcagctgctactttggggcggtatattat
141 L K A N M K A G T I G R A Y I P L T A A A T W A A Y Y
9268 cctttagggttgaagcatcatataacaaagtacaaaactatggttaaccatttttagacggtgcaataactattcttagcttgg
169 P L G L K A S Y N K V Q N Y G N P F L D G A N T I L A W
9352 ggtggttaaatagacggttaaggtggatcacctagtgattcgtctgacagtggttagtggtgacagtggttagtctactctc
197 G G K L D G K G G S P S D S S D S G S S G D S G S S L L
9436 gctttagcaaaaacacccatgcaagaatttataaaaaatacaagacgcattacaatgggacgttcatagtattggtagtgat
225 A L A K Q A M Q E L L K K I Q D A L Q W D V H S I G S D
9520 aaatttttttagtaaatgattattttacattagaaaaaacatttaacaacacatatcatattaaaaatgacgtattggtttacttgat
253 K F F S N D Y P T L E K T F N N T Y H I K M T I G L D
9604 tcattaaaaaaactgattgatagcgttcaagtagatagtgaggatagtagttctaactcctactgatgacggagaccataaa
281 S L K K L I D S V Q V D S G S S S S N P T D D D G D H K
9688 ccaattagtggttaaatcagtcagccaagtgaagaaagtggtcgtgtgattggtggttaactggacatgacacagttaccagaa
309 P I S G K S V K P N G K S G R V I G G N W T Y A Q L P E
9772 aaatataaaaaagcaattggtgtacctttattcaaaaaagaaactattatacaaacacaggttaacatatcttctcaaacgggtat
337 K Y K K A I G V P L F K K E Y L Y K P G N I F P Q T G N
9856 gcaggacaatgtacagaattaacatggcggtatgtcacaaactacatggttaaaagacaacctaccgacgacggtcaataaca
365 A G Q C T E L T W A Y M S Q L H G K R Q P T D D G Q I T
9940 aacggtcagcgtgtatggtacgtctataaaaagttagggtgcaaaaaacaacataatccaacagtaggttatgttctcctagt
393 N G Q R V W Y V Y K K L G A K T T H N P T V G Y G F S S
10024 aaaccaccatacttacaagcaactgcataatggttattggtcacacaggtggttggtagcagtttttagaagtggttcgttttta
421 K P P L Q A T A Y G I G H T G V V V A V F E D G S F L
10108 gttgcaaacataatgtaccaccatattgttcaccatcacgtgtggtattgtatcacactcattaatggcgtaaccaaatatgct
449 V A N Y N V P P Y V A P S R V V L Y T L I N G V P N N A
10192 ggtgataaatattgattctttagtggtattgcttaa 10227
477 G D N I V F P S G I A *
44AHJDORF005
13890 atggtaaaaaacaatcggttagacatggttaagagattatcaaaatgctgtcaatcatgtcagaaaaaaatcccgataagtat
1 M V K Q N R L D M V R D Y Q N A V N H V R K K I P D K Y
13806 aatcaaatagaattagttgatgaacrtatgaargatgatagattattatatactatttcaaacggtcttgatggaanaatcg
29 N Q I E L V D E L M N D I D Y Y I S I N R S D G K S
13722 ttcaactatggtttcattttttattttatttagctataaaacttgatataaaatttactttattatcacgtcattatcacattagct
57 F N Y V S F P I Y L A I K L D I K F T L L S R H Y T L R
13638 gacgcttaccgtgattttattgaagaaatcatagatgaaatccactatttaaatcaaaacggtgcacgttcagaagtgttagg
85 D A Y R D P I E E I I D E N P L F K S K R V T F R S A R
13554 gactatttagctattatctatcaagataaagaaattgggtgattacagatttgaaatagtgccactgatttaaatatcattct
113 D Y L A I I Y Q D K E I G V I T D L N S A T D L K Y H S
13470 aacttttttaaacactatcctattattatattatgagtttttagcacttgaagatgattatttaattgatgagtggtgataag
141 N F L K H Y P I I I Y D E F L A L E D D Y L I D E W D K
13386 ttaaaaaacaatatatgaatcaatcgacgttaaccatggttaacgttgattatattggattccctaaaaatgttttactaggtaat
169 L K T I Y E S I D R N H G N V D Y I G F P K M F L L G N
13302 gcagtcacacttttcaagtcctattatccaattttaatatatacaattttattacaaaagcataaaatgaatcacatcaagactt
197 A V N P S S P I L S N L N I Y N L L Q K H K M N T S R L
13218 taaaaaaacattttttgaaatgcgacgaacgattacgtgaatgaaaaacgtaacacacggtgcttttaattcaaatgacgac
225 Y K N I F L E M R R N D Y V N E K R N T R A P N S N D D
13134 gctatgacaactggagaatttgaatttaacgaatataatttggcggatgataatttaagaaatcacatcaatcaaaacggtgat
253 A M T T T G E F E F N E Y N L A D D N L R N H I N Q N G D
13050 ttcttctatatacaaaactgataataatataaaagtcataatgtaactacttttatgacaaaatattatcgtttgtatcca
281 F F Y I K T D D K Y I K V M Y N V T T P M T N I I V V P
12966 tatacaaaacaatatgaattttgtactaaaattaggatagacaatcatgttacctatttaccgtgatgatgttttataaa
309 Y T K Q Y E P C T K I R D I D N H V T Y L R D D M F Y K
12882 gaaaacatggaacggtattactacaatccaagcaattttacattttgacaatgcttacttcaaaaattacgtggttgatgaatgat
337 E N M E R Y Y Y N P S N L H F D N A Y S K N Y V D N D
12798 agatattttatattagatatgaataaaattttcatataaaaaatgaaatgaagaaaaatagagtgagtttgaaaga
365 R Y L Y L D M N K I I K P H I K N E M K K N M S E F E R
12714 aaagaaaaatatagcaagataaactatagagaatagcaaaaaggtatctaagaaacaatcggtctataa 12643
393 K E K I Y E D N Y I E N T K K Y L M K Q Y G L *
44AHJDORF006

803 atggcacaacaatctacaaaaatgaaactgcacttttagtagcaagtcagctaaatcagcggttacaagattttaatcatgat
1 M A Q Q S T K N E T A L L V A K S A K S A L Q D F N H D
887 cattcaaaatcttgacatttgccgacaaatgggataattcaaatcaaatgttcgaacattttgtaataaatatttattccct
29 Y S K S W T F G D K W D N S N T M F E T F V N K Y L P P
971 aagattaatgagactttattaatcgatattgcattaggaatcggttttaattggtagctaaagagcaagattttattggacaa
57 K I N E T L L I D I A L G N R F N W L A K E Q D F I G Q
1055 tatagtgagaataacgtgattatggacacagtaaccaatcaacatggacttatctaaaaatgaggaattaatgttgaaacgtaat
85 Y S E E Y V I M D T V P I N M D L S K N E E L M L K R N
1139 tatccacgtatggcaactaagttatatggtaacggaattgtgaagaacaaaaattcacattaaacaacaatgatacacggttc
113 Y P R M A T K L Y G N G I V K K Q K F T L N N N D T R F
1223 aattttccaaacattagcagacgcaactaattacgcttttaggtgtatcaaaaagaaaaatttctgatattaatgtattagaagaa
141 N F Q T L A D A T N Y A L G V Y K K K I S D I N V L E E
1307 aaagaatgcgtgcaatggttagttgattacttcaattcaattatccgaacaaatgtagctaaagcaacatcaaaagagat
169 K E M R A M L V D Y S L N Q L S E T N V R K A T S K E D
1391 ttgacagcaaaagtttttgaagcaactcctaaacttacaanaacacagtgctaaatataatgaagtagcatcggtgcatcaggcggt
197 L A S K V F E A I L N L Q N N S A K Y N E V H R A S G G
1475 gcaattggacaaatatacaactgtatcaaaataaaagattattgtgatttttaacacagattcattaaaaatcttatttttagat
225 A I G Q Y T T V S K L K D I V I L T T D S L K S Y L L D
1559 actaagattgcaaacacattccagattgcaggcattgattttcacagatcacgtttattagtttttgacgacttaggtggcggttt
253 T K I A N T F Q I A G I D F T D H V I S F D D L G G V F
1643 aaagtacaaaaagaatttaagttacaaaaccaagattcaattgactttttacgtgctgattgagattatcaatcacaattagga
281 K V T K E F K L Q N Q D S I D F L R A Y G D Y Q S Q L G
1727 gatacaattccagttggtggtgattttacttattgatgtatctaaacttaaaagagtttactggcaacgttgaagaattaaacca
309 D T I P V G A V F T Y D V S K L K E P T G N V E I K P
1811 aaatcagatttatatgcgtttatttttgatattaattcaattaaatataaaacttacacaaaagggtatgttaaacccaccattc
337 K S D L Y A F I L D I N S I K Y K R Y T K G M L K P P F
1895 cataacctgaatttgatgaagttacacattgatttacttacttttaaaagccattagtcattctttaaataaaatttta
365 H N P E F D E V T H W I H Y Y S F K A I S P F P N K I L
1979 attactgaccaagatgtaaatccaaaaccagaggaattacaagaataa 2029
393 I T D Q D V N P K P E E E L Q E *
44AHJDORF007
2044 atgaacaacgataaaagaggttttaacggttagttatcaaaaggaatcagcaaaagaggtgttgaacatcgcaacagattttaa
1 M N N D K R G L N V E L S K A E I S K R V V E H R N R P K
2128 cgttttatgttttaacgtttatttggaaattttaccgctactaatcaactataccaatcggtatcggttggtatagattttatt
29 R L M F N R Y L E F L P L L I N Y T N R D T V G I D F I
2212 cagttagaatcagcttttaagacaaaacattaatgtagttgttggtagaagtagaataagcaaatatgattcttgggttatgta
57 Q L E S A L R Q N I N V V V G E A R N K Q I M I L G Y V
2296 aataacacttactttaatcaagcacaatctttcatcaaaactttaatttccaaatttcaaaaacgataataaagaagata
85 N N T Y P N Q A P N F S S N P N P Q P Q K R L T K E D I
2380 tattttattgtacgtgactatttaaatctgtatgttctacaaattcataagcttatatgataactgtatgagtggttaacttt
113 Y F I V P D Y L I P D D C L Q I H K L Y D N C M S I N Q L
2464 gttgtcatgcaaaaataaaccaattcaatataatcagtgatatagaattatagaacattatactgatgaattcagagaaggtgt
141 V V M Q N K P I Q Y N S D I E I I E H Y T D E L A E V A
2548 ttactctgctttttttaaactcatgcaagcaaaatttagaagatatttaaatcagaataatgaacagtgcaatcaatcaactt
169 L S R F S L I M Q A K F S K I P K S E I N D E S I N Q L
2632 gtgtccgaaatatataacggtgcaccattgttaaaactgtccactatgtttaaagcagatgacgatatcattgtttaacaaag
197 V S E I Y N G A P F V K M S P M F N A D D D I I D L T S
2716 aatagcgttaatcccagcattaaactgaatgaacgggaatatacaaaaataattagtgaaattgaactatttaggcattaat
225 N S V I P A L T E M K R E Y Q N K I S E L S N Y L G I N
2800 tcattagcgttgataaagaagcggtgtttcagacgaagaggcaaaaagtaactcgtggattttaccacatcaaacagtaaatc
253 S L A V D K E S G V S D E E A K S N R G F T T S N S N I
2884 tatttaaaagggtcggtgaaccaattacgtttttatcaaaagcgttatggttttagatattaaaccgtattacgatgatgaacaacg
281 Y L K G R E P I T F L S K R Y G L D I K P Y Y D D E T T
2968 tctaaaatataatggttagacacactttttaagatgaagcagtgatataaatggctag 3027
309 S K I S M V D T L F K D E S S D I N G *
44AHJDORF008
3020 atggctagatacacaatgactttatcagattttcattaaatcagaattgattaaaaagggtttcaatgaattttgtaaatgataat
1 M A R Y T M T L Y D F I K S E L I K K G F N E F V N D N
3104 aaatttaacgttttatgatgatgaatttcaattcatgcaaaaatgctgaagttcgacaaagacgttttagctatcggttaatgaa
29 K L T F Y D D E F Q F M Q K M L K F D K D V L A I V N E
3188 aaagtatttaagggtttttcattgaagatgaattatcagattttacttttttaaaaaatcatttaccgattcatttttttagataga
57 K V F K G F S L K D E L S D L L F K K S F T I H F L D R
3272 gaaatcaacagacaaacagttgaagcattttggcagtgagtgattactgtatgtattacacatgaggattatttaaatgtggtt
85 E I N R Q T V E A F G M Q V I T V C I T H E D Y L N V V
3356 tatccatcaagtgaagttgaaaaatacttacaatcacaaggcttcacagaacacaaatgaagatacaacaagtaacactgatgaa
113 Y S S S E V E K Y L Q S Q G F T E H N E D T T S N D E
3440 acatcgaaatcaaaatgctacatctttagacaattcaactggcagtgactgcaaacagaaacgcttatgtgtcattaccacaaagt
141 T S N Q N A T S L D N S T G M T A N R N A Y V S L P Q S
3524 gaggttaacattgatgttgataatacaacgttaccgttgcgtgataataatcagattgataacggtaaaactgtgataaatcg
169 E V N I D V D N T T L R F A D N N T I D N G K T V N -K-S
3608 agtaacgaagtaatacaaaacgcaaaacgtaatacaaaatgcaaaaggtaataacgaaggtacacaaattcactaagcagatttta
197 S N E S N Q N A K R N Q K G N A K G T Q F T K Q Y L
3692 attgataatattgataaagcgtacgatttaagaagaagaaatttttaaatgaatttgataaaaaatggttttttacaattcggtag
3775
225 I D N I D K A Y D L R K K I L N E F D K K C F L Q I W *
44AHJDORF009

113 E S N K E D N D Y S D E E L V D K L D L D *
44AHJDORF114
16172 atgggttaattgtgataatgcaccagaagaaaaaggacaagccctatactgaaatgttgcaactattcaataaactgattcaatgg
1 M V N V D N A P E E K G Q A Y T E M L Q L F N K L I Q W
16088 aatccagctttatacatttgacaatgaacttattatcggttgccaacactattattaaactataatagttctgtgtt
29 N P A Y T F D N A I N L L S A C Q Q L L L N Y N S S V V
16004 caattcttaattgatgaactaaacaacgaactaaaccagaatcaatattgtcttatattgctggtgatgacccaatagaacaa
57 Q P L N D E L N N E T K P E S I L S Y I A G D D P I E Q
15920 tggaaatgcataaaggattttatgaaacgtataacgtttacgttttttag 15870
85 W N M H K G F Y E T Y N V Y V F *
44AHJDORF014
6243 atgaaaatgggtacatttacatgtggttttttaccatatttgacgtgtcggttagtccaaaattatcagaacctaatggctatt
1 M K M V H L H V V F Y Q Y L H V S V V Q N Y Q N L M A I
6327 gggttccaaacacgggttatcaccatataacgaagtttttatcagatgggtacgtatggattgggtataactggcgaaggca
29 G S N Q T V I H H I T K F V Y Q M V T Y G L V I T G K A
6411 cagcttattattaccagtgcccaatggaatggaacacaggttaattagttacagtggtgttattccttgggggggttctcat
57 H V I I Y Q C A N G M E K Q V I V T V L V F L G G C S H
6495 aatgggtatttttagcctttttcttga 6521
85 N G Y F S L F L *
44AHJDORF015
15403 gtgacgataaacacctgttcaccgaattttgattctttgtttgtgaataacgtcttaacgatatactctttttcataccgtat
1 V T I T P C S P N F D S L F V N N A L T I Y S P F I P Y
15487 tttttactaattctgatagtttgataaattctctttcttttctcctcaaatcaaatctcgtaattgtgtttgtgtcttgat
29 F S T N S D S L I N S L S F S S N L A N V P W C L D
15571 aaaaatctctttacgtttgtcattttattctctcttatttaattatttgccttctgcaattgcatgttag 15645
57 K I S F T F V I L F L L L F K L F A F C N C D L *
44AHJDORF016
15852 atgaaagtgtgacacattgttaccttactgttcaaggttatatacttacttacttagatgatgataatgaatacattgaggaa
1 M K V D D I V T L R V K G Y I L H Y L D D D N E Y I E E
15768 tttttaccatttcacgagtatcatttaacaaaacacagaagaattattaccagacacatgtaactattgtccactaca
29 F L P L H E Y H L T K T Q A K E L L P D T C K L L S T T
15684 cgcacaacgaaaacaattcaagtttattacaatgattttactacaatcgcaattgcagaaagcaataa 15616
57 R T T K T I Q V Y Y N D L L Q I A I A E S K *
44AHJDORF017
10757 atgaaaagattaaaattgcttctgtgtgtataccgaaaaacgcctttgatataacgctgattttgaaacctttgtacgtgaac
1 M E R L K L L L L V Y R K T P L I Q A S I L K P L Y V N
10673 aattctttgacggtgacattattgaaaacaataaagttatctataatgagcaaggtacaatatcgatatatccgtctaaaactg
29 N S L T V P L L K T I K V S I M S K V Q Y R Y I R L K L
10589 aaattgtatgtggtatgtatgatgaatattttactgatgaacttaatatga 10536
57 K L Y V V M Y M M N I L L M N L I *
44AHJDORF018
1098 atgttaattgggtactgtgtccataatcacgtattcttctactatattgtccaaataaaattctgtcttttagctaaccaattaaaa
1 M L I G T V S I I T Y S L Y C P I K S C S L A N Y I E E
1014 cgattacctaatgcaatatcgatttaataagttctcattaatcttagggaataaattattttacaaatgtttcgaaacttgta
29 R L P N A I S I N K V S L I L G N K Y L F T N V S N I V
930 cttgaattatcccatgtgtcccaaatgtccaagattttgaataa 886
57 F E L S H L S P N V Q D F E *
44AHJDORF019
9836 atgttacctggtttgtataagttattctttttgaataaaggtacaccaattgcttttttatatttttctggttaactgtgcatat
1 M L P G L Y K Y S F L N K G T P I A F L Y F S G N C A Y
9752 gtccagttaccaccaatcacacgaccactttttccatttggtgactgattaccactaattggtttatggtctccgtcatca
29 V Q L P P I T R P L P P F G L T D L P L I G L W S P S S
9668 tcagtaggattagaactactactccactatctacttga 9630
57 S V G L E L L L P L S T *
44AHJDORF121
16362 atggaaaatgaacaaaaaacattgagttgaagcatgtttttcgttttaagaatggaagtttatgtatagcgttatttgataga
1 M E N E T K N I E L K H V F R F K N G S L C I A L F D R
16278 acagaaaatgaattttcattttatgattgacattgaaattgaagatttaaatcataattctgttttacgcgttaatttca
29 T E N E I S F Y D V D I D E I E D L N H N S V L R V I S
16194 actttattaggaagtataataatggttaa 16165
57 T L L G S D N N G *
44AHJDORF020
13865 atgtctaaacgattttgttttaccatgtttttgctccttgtaaatgtttatgatgtcgtttacagtggttaaattttatcgctcaa
1 M S K R F C F T M F L L L V I V Y D V V Y S V K F I R Q
13949 atgttgcatataataaaagttataacctcacatcttcatcatcaatatttgcactggtctatctgatttaccattctttat
29 M L H N I K S Y T S H L H H Q Y L S L V Y L I Y Q F L Y
14033 ataaagtatcgattttcttttaa 14053
57 I K Y R P L *
44AHJDORF123
614 atgtatgagggaaacaacatgcgttctatgatgggtacatcatatgaagattcaagattaaataacgaacagaaattaaatgaa
1 M Y E G N N M R S M M G T S Y E D S R L N K R T E L N E
698 aacatgtcaattgatacaataaaagtggaagattgtatggtgtacaattcattcactttcaaaacaatcattacaggtgac
29 N M S I D T N K S E D S Y G V Q I H S L S K Q S P T G D
782 gttgaggaggaataa 796
57 V E E E *

44AHJDORF021

5816 atgcacccatcaaagtcaacacccgccccctcatgcttatatccattcttttgcttggtgttgatttcatttatcactc
1 M H H Q S Q H L P P H A Y I S I L L V V V I S F I S L
5732 ctatttttgatgttttgctacccaaccatattcacgatgtttgtttccgcattaacattactgaagaattctttatattccga
29 L F L M F C Y P T I F T M F C F R I N I T E E F F I F R
5648 tatattagcctcttaa 5634
57 Y I S L *

44AHJDORF022

8611 atgttttgctaaaatgataatacagaatatcaataatttttagaaaatcctctcattgatttttttgaccataagttattatt
1 M F A K M I I Q N I N N F L E N P L I D F F D H K L L F
8527 ttaattgcttttgaaatacctgtaataatatcaacgaacatttaatacaataaaaaagtag 8468
29 L I A F E I P V I I S T N I N T N K K *

44AHJDORF023

6494 atgagaacaccccccaaggaataacacactgtaactattacctgtttttccattccattggcgactggtaataataacgtg
1 M R T P P K E Y Q H C N Y Y L F F H S I G A L V N N N V
6410 tgccttgccagttataaccaatccatacgttaacatctgataaaacaaacttcgttatattggtgtataacgtttggtggaacc
29 C L A S Y N Q S I R N H L I N K L R Y M V Y N R L V G T
6326 aatagccattag 6315
57 N S H *

44AHJDORF024

14275 gtgtcaatgtacgcctcttgttaaatcttttatcatcaaattttaaaattacattactaaaatcatttaaaaaaatctttttct
1 V S M Y A S C K S L S S N L K L T L L K S F K N K S F S
14359 tgcctcttttctagcttctctttctttttccatctatccatttcagacgtatgtcttaaccaatgttatcaacctccatataaag
29 C S F L A S L S F F H L S I S D V C L T N V I N L H I K
14443 cataaataa 14451
57 H K *

44AHJDORF025

15175 atggaacgttaatacaaaaacggtattattatattgcgatgagattaaaggacattttccacatcaaatctcaatgtttgaagat
1 M E R K Y K T V L L Y C D E I K G H F P H Q I S M F E D
15091 ttatgtgacgttaagttgtatattcatattatgaatataacctgttcactaaaaaatacgcgtatatcatagaatacattaaag
29 L Y D A K V V Y S Y Y E Y N L P T K K Y A Y I I E Y I K
15007 gagatataa 14999
57 E I *

44AHJDORF026

14593 atgaataacctattataacatagccattgtttttcttttagcatttttaattacacttatcatacttatgacactgcatatacgc
1 M N N L L N I A I V F L L A F L I T L I I L M T H I R
14509 gtgtcatttgggtgttttattcactacattgatttatattctatattatcttttaagggttatttatgctttatattgagggtga
29 V S F G V L F T T L I I F Y I I F L M V I Y A L Y G G *

44AHJDORF027

12916 atgattgtctatattccctaatttttagtacaataatcatattgttttgtatattggtacaacgataatatttgcataaaaagtagt
1 M I V Y I P N F S T K F I L P C I W Y N D N I C H K S S
13000 tacattatacatgactttaatatatttatcatcagttttgtatagagaagaatcacgcgttttgattgatgtgatttcttaa
13080
29 Y I I H D F N I F I I S F D I E E I T V L I D V I S *

44AHJDORF029

15183 gtgttttaaatggaacgttaatacaaaaacggtattattatattgcgatgagattaaaggacattttccacatcaaatctcaatgt
1 V F K W N V N T K R Y Y Y I A M R L K D I F H I K S Q C
15099 ttgaagatttatatgacgctaaagttgtatattcatattatgaatataacctgttcactaaaaaatacgcgtatatcatag
15019
29 L K I Y M T L K L Y I H I M N I T C S L K N T R I S *

44AHJDORF028

9235 atggaatatatgcacgtccaattgtacctgctttcatatttttgcaaaatctgcattaccttttctttgtacgtcttctgtgta
1 M E Y M H V Q L Y L L S Y F L Q N L H Y L F F V R L V V
9151 caaagtggaagatgttacctgcgtcataccaagacggtgtccagctgttttgattgtgataactttcttgctatga 9071
29 Q S G R C Y L R H T K T V V Q L V L I V I L T F L L *

44AHJDORF030

14487 gtgaataaaaacaccaaataacacgcgttatatgcagtgatcataagtgataagtgtaattaaaaatgctaaaaggaaaacaatg
1 V N K T P N D T R I C S V I S M I S V I K N A K R K T M
14571 gctatgttttaagggtattcatggtcaatcactttccattatcgatatgactttgttttgataaataatcattaa 14648
29 A M P N R L F M V N H F P I I V Y D F V L I N N H *

44AHJDORF031

11039 atgatattgtatagttcattgttatcatctaaacggaataagttaaaatgtgaacgtaaatgcaggtatgccatataatccattt
1 M I L Y S S L L S S K R N K L K C E R N A G M P Y N P F
11123 aaaacgacttttagataacataacctcctcatttgagtaggggtgttcgttgatcatcagtaagtga 11191
29 K T T L D N I T S S F E Y G C S L I S S V M *

44AHJDORF135

693 atgaaaacatgtcaattgatataaaaagtgaagatagttatgggtgacaaaattcattcactttcaaaaacatcatttacag
1 M K T C Q L I Q I K V K I V M V Y K F I H F Q N N H L Q
777 gtgacgttgaggaggaataataaattatggcacaacaatctacaaaaaatgaaactgcacttttag 842
29 V T L R R N N K L W H N N L Q K M K L H F *

44AHJDORF033

3795 atgccattatttaaccacacctaccacaaatttgtaaaaaacattttttatcaaattcatttaaaattttctttcttaaatcgta
1 M P L F N H L Y Q I C K K H P L S N S F K I F F L K S Y

280

11995 atctttatcaaaaatcgataaataaaatctgttttaagttgtga 12039
29 I F I K N R I I K I C F K L *
44AHJDORF045
10655 atggcaccgtcaagaattgttcacgtacaaagggttcaaaatcgacgcttgatcaaaaggcggttttcgggtataccagcagaa
1 M A P S K N C S R T K V S K S T L V S K A F F G I P A E
10739 gcaattttaatctttccattcattcatatgcatatttcttatga 10783
29 A I L I F P F T S Y A Y F L *
44AHJDORF048
15340 atgaggacgtgtgttgacattatcaatgctggagaagtccaattcacaatttatgaatatgaaaacaaaaagggtcaaaagggt
1 M R T L L T L S M L E K F N S Q F M N M K T K K V K K V
15256 actcaatcaatttgggtcaagtatcattttaatacaatttcataag 15212
29 T Q S I L V K Y H F N T I S *
44AHJDORF049
5784 atgagggggcgaggtgttgactttgatggtgcatatggatttcaatgtatggacttatcagttgcttatgtgtattacattactg
1 M R G Q V L T L M V H M D F N V W T Y Q L L M C I T L L
5868 acggtaaagtctcgcatgtgggttaatgctaaagacgcgataa 5909
29 T V K F A C G V M L K T R *
44AHJDORF050
13158 gtgtgttacgtttttcattcacgtaatcgcttcgctgcatttctaaaaaatgttttctgaagtccttgatgtattcattttat
1 V C Y V F H S R N R F V A F L K K C F C K V L M Y S F Y
13242 gcttttgaataaattgtatatatttasattggataatag 13283
29 A F V I N C I Y L N W I I *
44AHJDORF051
11066 atgataacaatgaactatacaatattcattaacggttacaaaaacactgaacgtaatatattattctctacatttgcacatcac
1 M I T M N Y T I S L T V T K T L N V I Y Y S L H L S H H
10982 gttcattgtataacttattggttctcttccaatacttaa 10944
29 V H C I T Y W F L S N T *
44AHJDORF052
14338 atgatttttagtaattgttaatttttaatttgatgataaagatttacaagaggcggtacattgacacatggaacattttgcacatc
1 M I L V M L I L N L M I K I Y K R R T L T H G N I L H I
14254 tgccctatttttctaagaagaacgatacatatgtaa 14216
29 C P I F L K K E T Y H M *
44AHJDORF053
3348 atgtgggtttattcatcaagtgaagtgaataacttacaatcacaaggcttcacagaaacacaatgaagatacaacaagtaaca
1 M W F I H Q V K L K N T Y N H K A S Q N T M K I Q Q V T
3432 ctgatgaaacatcgaaatcaaatgctacatcttttag 3467
29 L M K H R I K M L H L *
44AHJDORF054
7551 atgactggaatggaataacgatgttactcgacgctggaagatttcacaaaaactgggtgtaagtacgtacaaaatcaatta
1 M T G M E I R C Y S T L V R F H K K L V L S Y V Q N Q L
7635 ttggttatcataatgaagttcgagtatatccagtag 7670
29 L V I I M K F E Y I Q *
44AHJDORF055
15705 atgtgtctggtgaataattcttttctgtgttttgggttaaatgatactcgtgaagtggttaaaattcctcaatgtattcattat
1 M C L V I I L L L V F W L N D T R E V V K I P Q C I H Y
15789 catcatctaagtaattgaagtataaacctttga 15821
29 H H L S N E V Y N L *
44AHJDORF056
5512 gtgagttattacattacaggttaaccaaatggaattatttttagagacgcccagaagaattaaaaaagggtgggtgcattgggtacgtg
1 V S I T L Q V T K W N Y L E T R Q K K L K K W V H G Y V
5596 tgtcaagtggttaacgcagtcggtgaagtaa 5625
29 C Q V V T Q S V K *
44AHJDORF057
10121 atgtaccaccatattgttcaccatcacgtgtgtgtattgtatacactcattaatggcgtaaccaataatgctgggtgataatattg
1 M Y H H M L H H H V W Y C I H S L M A Y Q I M L V I I L
10205 tattcttttagtggtattgtttaattaa 10231
29 Y S L V V L L N *
44AHJDORF058
10767 atgcatatttcttatgattcagttacaaacatcttatctatctgttcgttttcaatatccatttacctaaggctatcgggtcga
1 M H I S Y D S V Q T S Y L S V R F Q Y P I Y L R L S G R
10851 ataaactgggttcaataagggttttaa 10877
29 I N W G S I R V *
44AHJDORF164
702 atgttttcatttaattctgttcttttatttaattcttgaatcttcatatgatgtacccatcatagaacgcagtggtgttccctca
1 M F S F N S V R L P N L E S S Y D V P I I E R M L F P S
618 tacatgtttaaatctctcctaatactaa 592
29 Y M F K F L L I *
44AHJDORF059
8360 atggattttgtaacattggattacctgaaccgtcattatgccaaaatcttacaccagattctaaaaattgcttttaattgttcca
1 M D F V T L D Y L N R H Y A K I L H Q I L K L L L I V P
8276 ttaacatgggttcgatgtcacgtatag 8250
29 L T W G R C H V *
44AHJDORF060
6257 atgtaccattttcatttctataatattgtgcggtattgttctgtttccattttccaaatgtatttacttttgatgtttctaattg
1 M Y H F H F Y N M C R I G F V S I F Q M Y L L L M F L M

6173 ctttgctattactacctgaaaatttag 6147
29 L C Y Y Y L K I *
44AHJDORF061
15551 atggtgttttgggtccttgataaaaatcttttacgtttgtcatctttctcctcttatttaaattatttgctttctgcaatt
1 M C F G V L I K Y L L R L S P Y F S S Y L N Y L L S A I
15635 gcgattttagtagtaaatcattgtaa 15658
29 A I C S K S L *
44AHJDORF062
4285 gtggatttcgcaacgcagttaaccaatcttattaatattgataaagaacaaatcacatgtactctacacaatccgattctcaaa
1 V V F A T Q L T N L L I L I K K Q I T C T L H N P I L K
4369 aacctgaagggtttttggataa 4389
29 N L K V F G *
44AHJDORF063
9487 atggtgttctgtatttttttaataattcttgcattggcttgttttgctaaagcgagtagtgaaactaccactgtcaccactactac
1 M R L V F F L I I L A W L V L L K R V V N Y H C H H Y Y
9403 cactgtcagacgaatcactag 9383
29 H C Q T N H *
44AHJDORF065
5029 gtgggtggaaaacgtttccataatttatatgggtttcttccaacttggtgagtatgaacactttgaagcattacgcgcaagaggtt
1 V V E N V S I I Y M V S S N L V S M N T L K H Y A Q E V
5113 cacaaaactataaattaa 5130
29 H K T I N *
44AHJDORF064
2609 atgacgagtcgaatcaatcaacttgggtccgaaatatataacgggtgcaccatttgttaaaatgtcacctatgtttaatgcagatg
1 M T S Q S I N L C P K Y I T V H H L L K C H L C L M Q M
2693 acgatatcattgatttaa 2710
29 T I S L I *
44AHJDORF066
10481 atgatattctttatattgaaagtgcacatcggttcattttcacttaacgacttatttccagttgaacgttcagtacataacaaat
1 M I F F I L K V T S V H F H L T T Y F Q L N V Q Y I T N
10397 ctgatttgcataatattaa 10380
29 L I C I Y *

Table 19

Sequence similarities between ORFs 44AHJD and public databases

Phage: 44AHJD

Database: nr

Query= sid|110871|lan|44AHJDORF001 Phage 44AHJD ORF|10342-12627|-1
(761 letters)

gi 118848 sp F19894 DPOL_BPM2 DNA POLYMERASE >gi 76896 pir JQ0...	55	1e-06
gi 1072656 pir S51275 DNA polymerase - phage CP-1 >gi 836593 e...	53	6e-06
gi 1429230 emb CAA67649 (X99260) DNA polymerase [Bacteriophage...	49	1e-04
gi 1572479 emb CAA65712 (X96987) DNA polymerase [Bacteriophage...	46	0.001
gi 118851 sp P06950 DPOL_BPPZA DNA POLYMERASE (EARLY PROTEIN GP...	45	0.002
gi 2435429 AF012250 unassigned reading frame (possible DNA po...	45	0.002
gi 1084487 pir S41618 DNA polymerase - slime mold (Physarum po...	45	0.002
gi 4877819 gb AAD31446.1 (AF133505) DNA polymerase [Neurospora...	44	0.004
gi 461962 sp P33537 DPOM_NEUCR PROBABLE DNA POLYMERASE >gi 2833...	44	0.004
gi 2499511 sp Q12471 6P22_YEAST 6-PHOSPHOFRUCTO-2-KINASE 2 (PHO...	41	0.041
gi 2258375 gb AAD11909.1 (AF007261) transcription initiation f...	40	0.070
gi 15734 emb CAA37450 (X53370) DNA polymerase (AA 1-575) [Bact...	39	0.092

Query= sid|110872|lan|44AHJDORF002 Phage 44AHJD ORF|3789-5732|3
(647 letters)

gi 135273 sp P27622 TAGC_BACSU TEICHOIC ACID BIOSYNTHESIS PROTE...	112	7e-24
gi 142847 (M64050) DNase inhibitor [Bacillus subtilis]	52	1e-05
gi 4038407 (AF103943) factor C protein precursor [Streptomyces ...	39	0.10

Query= sid|110873|lan|44AHJDORF003 Phage 44AHJD ORF|6626-8389|2
(587 letters)

gi 138123 sp P04331 VG9_BPPH2 TAIL PROTEIN (LATE PROTEIN GP9) >...	92	8e-18
gi 138124 sp P07534 VG9_BPPZA TAIL PROTEIN (LATE PROTEIN GP9) >...	82	1e-14
gi 1429238 emb CAA67657 (X99260) tail protein [Bacteriophage B...	78	2e-13
gi 215339 (M12456) p9 tail protein [Bacteriophage phi-29] >gi 2...	71	2e-11
gi 1181968 emb CAA87738.1 (Z47794) tail protein [Bacteriophage...	54	3e-06
gi 1181970 emb CAA87740.1 (Z47794) tail protein [Bacteriophage...	42	0.010

Query= sid|110875|lan|44AHJDORF005 Phage 44AHJD ORF|12643-13890|-1
(415 letters)

gi 3845203 (AE001399) GAP domain protein (cyclic nt signal tran...	52	6e-06
gi 3758843 emb CAB11128.1 (Z98551) predicted using hexExon; MA...	49	5e-05
gi 3845297 (AE001421) hypothetical protein [Plasmodium falciparum]	48	1e-04
gi 4493936 emb CAB38972.1 (AL034556) predicted using hexExon; ...	47	2e-04
gi 3845165 (AE001390) hypothetical protein [Plasmodium falciparum]	46	6e-04

Query= sid|110877|lan|44AHJDORF007 Phage 44AHJD ORF|2044-3027|1
(327 letters)

gi 1181960 emb CAA87731.1 (Z47794) connector protein [Bacterio...	46	5e-04
gi 1429239 emb CAA67658 (X99260) upper collar protein [Bacteri...	45	8e-04
gi 137915 sp P07535 VG10_BPPZA UPPER COLLAR PROTEIN (CONNECTOR ...	44	0.002
gi 137914 sp P04332 VG10_BPPH2 UPPER COLLAR PROTEIN (CONNECTOR ...	41	0.009

Query= sid|110878|lan|44AHJDORF008 Phage 44AHJD ORF|3020-3775|2
(251 letters)

gi 4982468 gb AAD30963.2 (AF118151) SNF1/AMP-activated kinase ...	52	3e-06
gi 1730077 sp P18160 KYK1_DICDI NON-RECEPTOR TYROSINE KINASE SP...	46	2e-04
gi 3758855 emb CAB11140.1 (Z98551) predicted using hexExon; MA...	46	2e-04
gi 585795 sp P21538 REB1_YEAST DNA-BINDING PROTEIN REB1 (QBP) >...	46	3e-04
gi 172372 (M58728) DNA-binding protein [Saccharomyces cerevisiae]	46	3e-04
gi 2952545 (AF051898) coronin binding protein [Dictyostelium di...	45	6e-04
gi 535260 emb CAA82996 (Z30339) STARP antigen [Plasmodium reic...	45	7e-04
gi 1429240 emb CAA67659 (X99260) lower collar protein [Bacteri...	44	0.001

Query= sid|110879|lan|44AHJDORF009 Phage 44AHJD ORF|5744-6496|2
(250 letters)

gi 2764981 emb CAA69021.1 (Y07739) N-acetylmuramoyl-L-alanine ...	180	1e-44
gi 113675 sp P24556 ALYS_STAAU AUTOLYSIN (N-ACETYLMURAMOYL-L-AL...	118	6e-26
gi 1763243 (U72397) amidase (bacteriophage 80 alpha)	118	6e-26
gi 4574237 gb AAD23962.1 AF106851_1 (AF106851) LytN (Staphyloco...	84	9e-16
gi 3767593 dbj BAA33856.1 (AB015195) LytN (Staphylococcus aureus)	84	9e-16
gi 2764983 emb CAA69022.1 (Y07740) cell wall hydrolase Ply187 ...	77	2e-13
gi 3287732 sp O05156 ALE1_STACP GLYCYL-GLYCINE ENDOPEPTIDASE AL...	73	2e-12
gi 79926 pir A25881 lysostaphin precursor - Staphylococcus sim...	69	3e-11
gi 126496 sp P10548 LSTP_STAS LYSOSTAPHIN PRECURSOR (GLYCYL-GL...	69	3e-11
gi 3287967 sp P10547 LSTP_STASI LYSOSTAPHIN PRECURSOR (GLYCYL-G...	69	3e-11
gi 3341932 dbj BAA31898.1 (AB009866) amidase (peptidoglycan hy...	68	6e-11

Query= sid|110882|lan|44AHJDORF012 Phage 44AHJD ORF|8391-8813|3
(140 letters)

gi 140528 sp P24811 YQXH_BACSU HYPOTHETICAL 15.7 KD PROTEIN IN ...	80	6e-15
gi 4126631 dbj BAA36651.1 (AB016282) ORF45 (bacteriophage phi-...	76	1e-13
gi 141088 sp P26835 YNGD_CLOPE HYPOTHETICAL 14.9 KD PROTEIN IN ...	61	4e-09
gi 2293160 (AF008220) YtkC (Bacillus subtilis) >gi 2635548 emb ...	36	0.099
gi 1181973 emb CAA87743.1 (Z47794) holin protein (Bacteriophag...	31	3.3

Table 20

Homologies between phage 44 AHJD ORFs and proteins in public databases

Query= pt|110871 44AHJDORF001 Phage 44AHJD ORF |10342-12627|-1 1
(761 letters)

>gi|118848|sp|P19894|DPOL_BPM2 DNA POLYMERASE >gi|76896|pir||JQ0161
DNA-directed DNA polymerase (EC 2.7.7.7) - phage M2
>gi|215509 (M33144) DNA polymerase (Bacteriophage M2)
Length = 572

Score = 55.4 bits (131), Expect = 1e-06
Identities = 96/426 (22%), Positives = 159/426 (36%), Gaps = 88/426 (20%)

Query: 229 KLTPEQLTYIHNDVILGMCHIHYSDFPNFDYNKLTFSNLNIMESYLNEMTR-----FQ 283
++TPE+ YI ND+ I+ DI +++T + + + + T+ F
Sbjct: 154 EITPEEYIKNIDIBIARA-----LDIQFKQGLDRMTAGSDSLKGFIDILSTRKFNKVFP 209

Query: 284 LLNQYQDIKISYTHYHFDNMNFDYIKSFYRGGLNMYNTKYINKLIDEPCEFSIDINSSYP 343
L+ D +I + YRGG N KY K I E D+NS YP
Sbjct: 210 KLSLPMDEKI-----RKAYRGGFTWLNDKYKEKEIGGMV-FDVNSLYP 252

Query: 344 YVMYHEKIPTWLYFYEHYSEPTLIPTFLDDDNYSFLYKIDKDVFNDDLLIKISRVLRQM 403
MY +P Y P + + D + LY I + P +L K + +
Sbjct: 253 SQMYSRPLP-----YGAPIVFQGYEKDEQYPLY-IQIRPEFEL---KEGYIPTI 299

Query: 404 XXXXXXXXXXXXXXXXXXXXLRMIQ-DITGIDCMHVRNSFVIYECYPHARDIIFQNYFIK 462
+ ++ +T +D I+ + + +Y EY F +
Sbjct: 300 QIKKNPFKNGEYLNKSGVEPVELYLTNVDELLEIHEH-YELYNVEYIDGFK-----FRE 352

Query: 463 TQGLKKNKINMTSPDYHITDDINEHPYSNEEVMLSKVVLNGLYG-----IPAL 511
G K+ I+ + H + L+K++LN LYG +P L
Sbjct: 353 KTLGFKDFIDKWTYVKTH-----BEGAKKQLAKMLNLSLYGKFASNPVDTGKVPYL 403

Query: 512 RSHFNL-FRLDDNNELNYIINGYKNTERNILFSTFVTSRSLYNLLVPPQYLTSBIDDNF 570
+ +L FR+ D YK+ + F+T+ + + + Q D
Sbjct: 404 KDDGSLGFRVGDDE-----YKDPVYTPM-GVFITAWARFTTITAAQACY-----DRI 449

Query: 571 IYCDTDSLYMKSVVKPLNPSLFDPIALGKWDIENEQIDKMPVLNHHK-----YAYEVNG 625
IYCDTDS+++ P + + DP LG W E+ + L K Y EV+G
Sbjct: 450 IYCDTDSIHLTGTEVPEIINKDIVDPKGLGYAHES-TPKRAKYLRQKTYIQDIYVKEVDG 508

Query: 626 KIKIAS 631
K+K S
Sbjct: 509 KLKECS 514

>gi|1072656|pir||S51275 DNA polymerase - phage CP-1
>gi|836593|emb|CAA87725.1| (247794) DNA polymerase
(Bacteriophage CP-1)
Length = 568

Score = 53.5 bits (126), Expect = 6e-06
Identities = 104/464 (22%), Positives = 169/464 (36%), Gaps = 66/464 (14%)

Query: 230 LTPEQLTYIHNDVIL--GMCHIHYSDFPNFDYNKLTFSNLNIMESYLNEMTRFQLLNQ 287
+ PE + YIH DV IL G+ ++Y + P Y + +L + +F+
Sbjct: 152 IKPEWIDYIHVDVAILARGIFAMYRENFTEK--YTSASEALTEFKRIFRKSRRKFRDFPP 209

Query: 288 YQDIKISYTHYHFDNMNFDYIKSFYRGGLNMYNTKYINKLIDEPCEFSIDINSSYPVVMY 347
D K+ D+ + G + K+ + +++ DINS YP M
Sbjct: 210 ILDEKVD-----DPCRKHIVGAGRLPTLKHGRGRTNLQIDIDIYDINSMPATML 257

Query: 348 HEKIPTWLYFYEHYSEPTLIPTFLDDDNYSFLY-KIDKDVFNDDL-LIKISRVLRQMXX 405
+P + + Y P + +D+Y+ + K D D+ L I+IK ++
Sbjct: 258 QNALPIGIP--KRYKGG---PKEIKEDHYIYHIKADFDLKRGLPTIQIKKLDALRIG 312

Query: 406 XXXXXXXXXXXXXXXXXXXXLRMIQDITGIDCMHVRNSFVIYECYPHARDIIFQNYFIK 465
L + + H + E F +F +Y
Sbjct: 313 VRTSDYVTTSKNEVIDLYLTNFDLFLKHGYDATIMYVETLE-FQTESDLFDDYI----- 366

285

Query: 466 KLQVKNMTSPYDYHITDDINEHPYSNEEVMLSKVVLNGLYGIPALR--SHFNLFRLLDDN 523
+ Y Y E+ S E +K++LN LYG + S L LDD
Sbjct: 367 -----TTYRYK-----KENAQSPAQKQKAKIMLSLYGKFGAKIISVKKLAYLDDK 412

Query: 524 NELNYNIINGYKNTERNIL-----FSTFVTSRSLYNLLVPFQYLTESEIDDNFIYCDTDS 577
L +KN + + FVTS + + + Q E DNF+Y DTDS
Sbjct: 413 GILR-----FKNDDEEEVQPVYAPVALFVTSIARHFIISNAQ-----ENYDNFLYADTDS 462

Query: 578 LYMKSVVKPLNPSLFDPIALGKWDIENEQIDKMFVLNHHKKYAYEVNGKIKIASAGIPKN 637
L++ +L+ DP GKW E + K L K Y E+ + + K
Sbjct: 463 LHLFMSDSLVLDD---IDPSEFGKWAHEGRAV-KAKYLRSLYIEELIQEDGTTTHLDV-KG 517

Query: 638 AFDTSVDFETVREQFFDGAIIENNKSIYNEQGTISIYPSKTEI 681
A T E E F GA E ++ +G IY + +I
Sbjct: 518 AGMTPEIKEKITFENFVIGATFEGKRASKQIKGGTLIYETTFKI 561

>gi|1429230|emb|CAA67649| (X99260) DNA polymerase [Bacteriophage
B103]
Length = 572

Score = 49.2 bits (115), Expect = 1e-04
Identities = 93/422 (22%), Positives = 155/422 (36%), Gaps = 88/422 (20%)

Query: 229 KLTPEQLTYIHNDVIIIGMCHIHYSDFPNFDYNKLTFSLNIMESYLNEMTR-----FQ 283
++TPE+ YI ND+ I+ DI +++T + + + + T+ F
Sbjct: 154 EITPEEYKYNKDIEIARA-----LDIQFKQGLDRMTAGSDSLKGFKDILSTKKFNKVFP 209

Query: 284 LLNQYQDIKISYTHYHFDNMNFYDIKSFYRGGLNMYNTKYINKLIDEPFCFSIDINSSYP 343
L+ D +I + YRGG N KY K I E D+NS YP
Sbjct: 210 KLSLPMDEKEI-----RRAYRGGFTWLDKYKEKEIGEGMV-PDVNSLYP 252

Query: 344 YVMYHEKIPTWLYFYEHYSEPTLIPTFLDDNNYFSLYKIDKDVFNDDLLIKIKSRVLRQM 403
MY +P Y P + + D + LY I + F +L K + +
Sbjct: 253 SQMYSRPLP-----YGAPIVFGQKYEKDEQYPLY-IQRIRFEFEL-----KEGYIPTI 299

Query: 404 XXXXXXXXXXXXXXXXXXXXLRMIQ-DITGIDCMHVRNSFVIYECEYFHARDIIFQNYFIK 462
++ +T +D I+ + + +Y EY F +
Sbjct: 300 QIKQNPFFKQNEYLKNSGASPEVLYLTNVDLELIQEH-YEMYNVEYIDGPK-----FRE 352

Query: 463 TQGLKQNKINMTSPYDYHITDDINEHPYSNEEVMLSKVVLNGLYG-----IPAL 511
G K I+ + H + L+K++ +LYG +P L
Sbjct: 353 KTGFLKFEIDKWTYVKTH-----EKGAKQLAKLMPDSLYGKFPASNPDTGKVPYL 403

Query: 512 RSHFNL-FRLDDNNELNYNIINGYKNTERNILFSTFVTSRSLYNLLVPFQYLTESEIDDNF 570
+ +L FR+ D YK+ + F+T+ + + + Q D
Sbjct: 404 KEDGSLGFRVGDDE-----YKDPVYTPM-GVFITAWARFTTITAAQACY-----DRI 449

Query: 571 IYCDTDSLYMKSVVKPLNPSLFDPIALGKWDIENEQIDKMFVLNHHK-----YAYEVNG 625
IYCDTDS+++ P + + DP LG W E+ + L K YA EV+G
Sbjct: 450 IYCDTDSIHLTGTEVFPEIHKDIDVPKILGYWAHES-TFKRAKYLKQKTYIQDIYAKEVDG 508

Query: 626 KI 627
K+
Sbjct: 509 KL 510

>gi|1572479|emb|CAA65712| (X96987) DNA polymerase [Bacteriophage
GA-1]
Length = 578

Score = 46.1 bits (107), Expect = 0.001
Identities = 80/376 (21%), Positives = 146/376 (38%), Gaps = 54/376 (14%)

Query: 234 QLTYYIHNDVIIIGMCHIHYSDFPNFDYNKLTFSLNIMESYLNEMTRFQLLNQYQDIKI 293
++ Y+ +D++I+ + +F N D+ +T + + +Y EM + +Y +
Sbjct: 162 EIEYLNKHDLLIVALA---LRSMFDN-DFTSMTVGSDALNTY--KEMLGVKQWEKYFFVL- 214

Query: 294 SYTHYHFDNMNFYDIKSFYRGGLNMYNTKYINKLIDEPFCFSIDINSSYPYVMYHEKIPT 353
+ I+ Y+GG N KY + + D+NS YP +M ++ +P
Sbjct: 215 -----SLKVNSEIRKAYKGGFTWVNPKYQGETVYGGMV-PDVNSMYPAMMKVLLP- 264

Query: 354 WLYFYEHYSEPTLIPTFLDDNNYFSLYKIDKDVFNDDLLIKIKSRVLRQMXXXXXXXXX 413
Y EP + + + LY F + KI ++

286

Sbjct: 265 -----YGEPMVFKGEYKKNVEYPLYIQVRCFFELKKDKIPCIQIKGNARFGQNEYLS 317

Query: 414 XXXXXXXXLRMIQDITGIDCMHIRVNSFVIYECEYFHARDIIFQNYFIKTQGKLKKNKINM 473
 L +T +D I+ + + I+E E+ +F+ + I

Sbjct: 318 TSGDEYVDLY----VTNVDELKIKH-YDIFEEFIGG--FMFKGF-----IGF 359

Query: 474 TSPYDYHITDDINEHPYSNEEVMLSKVVLNGLYGIPALRSHFN--LFRLDNNELNYINI 531
 Y + N S E+ + +K++LN LYG A + LD+N L

Sbjct: 360 FDEYIDRFMEIKNSPDSSAEQSLQAKMLNSLYGKFATNPDITGKVPYLDENGVLKFRKG 419

Query: 532 GYKINTERNILFST---FVTSRSLYNLLVPFQYLTESEIDDNFIYCDTDSLYMKS VVKPLL 588
 K ER+ +++ F+T+ + N+L Q L FIY DTDS++++ + +

Sbjct: 420 ELK--ERDPVYTPMGCFITAYARENILSNAQKLYP-----RPIYADTDSIHVEGLGEVDA 472

Query: 589 NPSLFDPIALGKWDIE 604
 + DP LG WD E

Sbjct: 473 IKDVIDPKKLG YWDHE 488

>gi|118851|sp|P06950|DPOL BPPZA DNA POLYMERASE (EARLY PROTEIN GP2)
 >gi|75812|pir|ERBP2Z DNA-directed DNA polymerase (EC
 2.7.7.7) - phage PZA >gi|216051 (M11813) gene 2 product
 [Bacteriophage PZA] >gi|224741|prf||1112171E ORF 2
 [Bacteriophage PZA]
 Length = 572

Score = 45.3 bits (105), Expect = 0.002
 Identities = 98/461 (21%), Positives = 166/461 (35%), Gaps = 110/461 (23%)

Query: 198 QLKTDPNYITFDKDNMDNDSEAYDYAVKCFAKLTPEQLTYIHNDVILGMCHIHYSDFP 257
 ++ DF T+ D D + Y ++TP++ YI ND+ I+ + I

Sbjct: 129 KIAKDFKLTVLKGDIDYHKERPVG Y-----EITPDEYAYIKNDIQAIEALL----IQF 178

Query: 258 NFDYNKLTFSLNIMESYLNEMTR-----FQLLNQYQDIKISYTHYHFDNMVFDYIKSF 312
 +++T + ++ + + T+ F L+ D ++ Y

Sbjct: 179 KQGLDRMTAGSDDLKGFKDIITTKFKKVPFPLSLGLDKVRYA----- 222

Query: 313 YRGGNMYNTKYINKLIDEPCEFSIDINSSYPYMYHEKIPTWLYFYZHYSEPTLIPT--P 370
 YRGG N ++ K I E D+NS YP MY +P Y EP +

Sbjct: 223 YRGGFTWLNDRFKEKEIGGMV-FDVNSLYPAQMYSRLLP-----YGEPIVFEGKYV 273

Query: 371 LDDNYFSLYKID-----KDVFNDDLKIKSRVLRQXXXXXXXXXXXXXXXXXXXXXLRMI 425
 D+D + I K+ + + IK +SR +

Sbjct: 274 WDEDYPLHIQHIRCEFEKELGYIPTIQIK-RSRFYKGYNEYLKSSGGEIADLW----- 324

Query: 426 QDITGIDCMHIRVNSFVIYECEYFHARDIIFQNYFIKTQGKLKKNKINMTSPYDYHITDDI 485
 ++ +D + + + +Y EY F T G K+ I+ + I

Sbjct: 325 --VSNVD--LELAKEHYDLYNVEYISGLK-----FKATTGLFKDFIDKWTIKITSEGA 375

Query: 486 NEHPYSNEEVMLSKVVLNGLYG-----IPALRSHFN--LFRLDNNELNYINIYG 533
 + L+K++LN LYG +P L+ + L FRL G

Sbjct: 376 KQ-----LAKMLNSLYGKFASNPVDTGKVPYKENGALGRL-----GE 415

Query: 534 KNTERNIL--PSTFVTSRSLYNLLVPFQYLTESEIDDNFIYCDTDSLYMKS VVKPLLNP 591
 + T+ + F+T+ + Y + Q D IYCDTDS+++ P +

Sbjct: 416 EETKDPVYTPMGVFITAWARYTTITAAQACF-----DRIIYCDTDSIHLTGTEIPDVIKD 470

Query: 592 LFDPIALGKWDIENEQIDKMFVLNHHKYAY-----EVNGKI 627
 + DP LG N E+ + L K Y EV+GK+

Sbjct: 471 IVDPKKLGYNHES-TFKRAKYLRQKTYIQDIYMKEVDGKL 510

>gi|2435429 (AF012250) unassigned reading frame (possible DNA
 polymerase) [Physarum polycephalum]
 Length = 544

Score = 44.9 bits (104), Expect = 0.002
 Identities = 118/545 (21%), Positives = 206/545 (37%), Gaps = 104/545 (19%)

Query: 179 TSIATLGKXLLDGGYLTESQLKTDPNYITFDKDNMDNDSEAYDYAVKCFAKLTPEQLTYI 238
 T + L K L D + T Q F N M Y + CF L P++ I

Sbjct: 62 TQLFNLLKSLQDSSPYTFKQ-----FTYQNM-----YSLEISCF--LYPKKKILI 105

Query: 239 HNDVILGMCHIHYSDFPNFD-----YNKL--TFSLNIMESY-LNNEMTRFQLLNQYQD 290
 D+ +I Y+D+ ++ YN++ +++NI Y L+ ++ +

Sbjct: 106 -KDLNFFSENIYNDVVKDYKLLAILYNEIQTAYNININRKYILSTASLSLRIFKKSFP 164

Query: 291 IKISYTHYHFDNMNFYDIKSFYRGGLNMYNTKYINKLIDPCFSIDINSSYPYVMYHEK 350
 K + D + +YI+ Y GG N I + + + + D+NS YPY+M EK
 Sbjct: 165 EKYRLIPHILTRDED--NYIRKSYIGGRNE-----IFEKVAQRNYFYDVNSLYPYIMKKEK 217

Query: 351 IPTWLYFYEHYSEPTLIPTFLDD-DNYFS----LYKIDKDVFNDDLL---IKIKSRVLQ 402
 +P + Y + + F + +N+F L I+K N +L + IK+ V
 Sbjct: 218 MPIGI---PEYRDKEYMKKFEKNIENFFGFIDVLITIEKTNNNIPVLPYRMGIKNV-EV 273

Query: 403 MXXXXXXXXXXXXXXXXXLRMIQDITGIDCMHIRVNSFVIYECEYPHARDIIFQNYFIK 462
 L + Q I+ IY + +++++F+ Y +
 Sbjct: 274 GIIYAKGTLRGIYFSEEIKLALKQGYKIIE-----IYSAYEYKEKEVVFEEYVEQ 323

Query: 463 TQOK-LQNKINMTSPYDYHITDDINEHPYSNEEVMLSQVVLNGLYG-----IPALRS 513
 + LK K D + D L K +LN LYG I +
 Sbjct: 324 MYNRRLKAK-----DPALKD-----LYKLLNTLYGRFGLVYEQIDIISP 363

Query: 514 HFNLFRLDDNNELNYNIINGYKNTERNILFSTFVTSRSLYNLLVPFQYLTESEIDNFIYC 573
 L + DN + + + + N + + + + F Y T + + IY
 Sbjct: 364 EKEL--ITDNTYISHDTTEFIDITANTCYNNIAITSAITSYARIFMYNTILNVLNHLVYI 421

Query: 574 DTDLSLYMKSQVVKPLNPSLFDPIALGKWDIENEQIDKMFVLNKKYAY-EVNGKIKIASA 632
 DTD L++K+ P+ + +L +GK+ +E+ + F+ N K Y Y +N I
 Sbjct: 422 DTDGLFLKN---PIPDIALTTSKEMGKFRLESINAEAHFIAN-KFYIYAPINSPIIYKFK 477

Query: 633 GIPK-----NAFDTSDVFETFVR----EQFFDGAIENNKSIYNEQGT-----ISIYPSK 678
 GIP N D + + +F +I NN Y+ Q + I Y +
 Sbjct: 478 GIPLQKPIFIHDIITQHKILNITLGHYFTFSIRLNNTQYSFQASRKRLIPNYKTT 537

Query: 679 TEIVC 683
 I+C
 Sbjct: 538 PWIIC 542

>gi|1084487|pir||S41618 DNA polymerase - slime mold (Physarum
 polycephalum) >gi|509721|dbj|BAA06121.1| (D29637) DNA
 polymerase [Physarum polycephalum]
 Length = 547

Score = 44.9 bits (104), Expect = 0.002

Identities = 118/545 (21%), Positives = 206/545 (37%), Gaps = 104/545 (19%)

Query: 179 TSIATLGKLLDGGYLTESQLKIDFNITFDKNDMDNDSEAYDAVKCFAKLTPEQLTYI 238
 T + L K L D + T Q F N M Y + CF L P++ I
 Sbjct: 65 TQLFNLLKSLQDSSFTPKQ-----PTYQNM----YSLEISCF--LYPKKKILI 108

Query: 239 HNDVILGMCHIHYSIDIPNFD-----YNKL--TFSLNIMESY-LANNEMTRFQLLNQYQD 290
 D+ +I Y+D+ ++ YN++ +++NI Y L+ ++ +
 Sbjct: 109 -KOLYNFFSENIIYNQVVKDYKLLAILYNEIQATYNNINRKYILSTASLSLRFKKSFP 167

Query: 291 IKISYTHYHFDNMNFYDIKSFYRGGLNMYNTKYINKLIDPCFSIDINSSYPYVMYHEK 350
 K + D + +YI+ Y GG N I + + + + D+NS YPY+M EK
 Sbjct: 168 EKYRLIPHILTRDED--NYIRKSYIGGRNE-----IFEKVAQRNYFYDVNSLYPYIMKKEK 220

Query: 351 IPTWLYFYEHYSEPTLIPTFLDD-DNYFS----LYKIDKDVFNDDLL---IKIKSRVLQ 402
 +P + Y + + F + +N+F L I+K N +L + IK+ V
 Sbjct: 221 MPIGI---PEYRDKEYMKKFEKNIENFFGFIDVLITIEKTNNNIPVLPYRMGIKNV-EV 276

Query: 403 MXXXXXXXXXXXXXXXXXLRMIQDITGIDCMHIRVNSFVIYECEYPHARDIIFQNYFIK 462
 L + Q I+ IY + +++++F+ Y +
 Sbjct: 277 GIIYAKGTLRGIYFSEEIKLALKQGYKIIE-----IYSAYEYKEKEVVFEEYVEQ 326

Query: 463 TQOK-LQNKINMTSPYDYHITDDINEHPYSNEEVMLSQVVLNGLYG-----IPALRS 513
 + LK K D + D L K +LN LYG I +
 Sbjct: 327 MYNRRLKAK-----DPALKD-----LYKLLNTLYGRFGLVYEQIDIISP 366

Query: 514 HFNLFRLDDNNELNYNIINGYKNTERNILFSTFVTSRSLYNLLVPFQYLTESEIDNFIYC 573
 L + DN + + + + N + + + + F Y T + + IY
 Sbjct: 367 EKEL--ITDNTYISHDTTEFIDITANTCYNNIAITSAITSYARIFMYNTILNVLNHLVYI 424

Query: 574 DTDLSLYMKSQVVKPLNPSLFDPIALGKWDIENEQIDKMFVLNKKYAY-EVNGKIKIASA 632
 DTD L++K+ P+ + +L +GK+ +E+ + F+ N K Y Y +N I
 Sbjct: 425 DTDGLFLKN---PIPDIALTTSKEMGKFRLESINAEAHFIAN-KFYIYAPINSPIIYKFK 480

Query: 633 GIPK-----NAFDTSDVFETFVR----EQFFDGAIENNKSIYNEQGT-----ISIYPSK 678

288

GIP N D + + +F +I NN Y+ Q + I Y +
 Sbjct: 481 GIPLQKPIFNHDIITQHKKILNITLGHYFTFSIRLNNNTYSFQASRKRKLIPNYKIT 540

Query: 679 TEIVC 683

I+C

Sbjct: 541 PWIIC 545

>gi|4877819|gb|AAD31446.1| (AF133505) DNA polymerase [Neurospora
 crassa]
 Length = 1035

Score = 44.1 bits (102), Expect = 0.004
 Identities = 36/172 (20%), Positives = 82/172 (46%), Gaps = 14/172 (8%)

Query: 521 DDNNELYNIIINGYKNTERNILFSTFVTSRSLYNLLVFPQYLTESEIDDNFIYCDTDSLVM 580
 + N EL + ++G K+ I ++ + + + + + S Y DTDS+++

Sbjct: 817 EKNYELLSYLDGEKDDGFIINSTSIAAATASWSRILMYKHIINSA-----YTDTSIFV 870

Query: 581 KSVVKPLLNPFLDPIALGKWDIENEQIDKMFVLNKKYAYEVNGKIKIASAGIPKNAFD 640
 + KPL + + + K + + I + ++ K Y + GK++I GI KN +

Sbjct: 871 E---KPLDSAFIGEGCGKFAEYNGQLIKRAIFISGKLYLLDFGGKLEIKCKGITKNKDN 927

Query: 641 TSVDFETFFVREQFFDG---AIIENNKSIYNEQGTISIYPSKTEIVCGNVYDE 689
 T+ + + E ++G + + E GT+++ K ++ G YD+

Sbjct: 928 TTHNLDINDFEALYNGESRVLFQERWGRSLELGTVTVKYQKYNLISG--YDK 977

>gi|461962|sp|P33537|DPOM_NEUCR PROBABLE DNA POLYMERASE
 >gi|283351|pir||S26985 probable DNA-directed DNA
 polymerase (EC 2.7.7.7) - Neurospora crassa
 mitochondrion plasmid maranhar (SGC3)
 >gi|578156|emb|CAA39046| (X55361) putative DNA
 polymerase [Neurospora crassa]
 Length = 1021

Score = 44.1 bits (102), Expect = 0.004
 Identities = 36/172 (20%), Positives = 82/172 (46%), Gaps = 14/172 (8%)

Query: 521 DDNNELYNIIINGYKNTERNILFSTFVTSRSLYNLLVFPQYLTESEIDDNFIYCDTDSLVM 580
 + N EL + ++G K+ I ++ + + + + + S Y DTDS+++

Sbjct: 815 EKNYELLSYLDGEKDDGFIINSTSIAAATASWSRILMYKHIINSA-----YTDTSIFV 868

Query: 581 KSVVKPLLNPFLDPIALGKWDIENEQIDKMFVLNKKYAYEVNGKIKIASAGIPKNAFD 640
 + KPL + + + K + + I + ++ K Y + GK++I GI KN +

Sbjct: 869 E---KPLDSAFIGEGCGKFAEYNGQLIKRAIFISGKLYLLDFGGKLEIKCKGITKNKDN 925

Query: 641 TSVDFETFFVREQFFDG---AIIENNKSIYNEQGTISIYPSKTEIVCGNVYDE 689
 T+ + + E ++G + + E GT+++ K ++ G YD+

Sbjct: 926 TTHNLDINDFEALYNGESRVLFQERWGRSLELGTVTVKYQKYNLISG--YDK 975

>gi|2499511|sp|Q12471|6P22_YEAST 6-PHOSPHOFRUCTO-2-KINASE 2
 (PHOSPHOFRUCTOKINASE 2 II) (6PF-2-K 2)
 >gi|2131162|pir||S61066 6-phosphofructo-2-kinase (EC
 2.7.1.105) - yeast (Saccharomyces cerevisiae)
 >gi|2131163|pir||S71026 6-phosphofructo-2-kinase (EC
 2.7.1.105) - yeast (Saccharomyces cerevisiae)
 >gi|1085116|emb|CAA62371| (X90861)
 6-phosphofructo-2-kinase [Saccharomyces cerevisiae]
 >gi|1420028|emb|CAA99157| (Z74878) ORF YOL136c
 [Saccharomyces cerevisiae] >gi|1628439|emb|CAA64733|
 (X95465) 6-phosphofructo-2-kinase [Saccharomyces
 cerevisiae]
 Length = 397

Score = 40.6 bits (93), Expect = 0.041
 Identities = 48/208 (23%), Positives = 92/208 (44%), Gaps = 29/208 (13%)

Query: 175 MKTNTSIATLGKKLLDGGYLTESQLKTDNFYITFDKDNMDNDSEAYDYAVKCFAKLTPEQ 234
 ++ S AT+ K LL L+ + + FN K+ND ++ +A++T ++

Sbjct: 139 IRRQISCATISKPLL---LSNTSSSEDLFN----PKNNDKKT-----YARITLQK 181

Query: 235 LTY-IHNDVIIIGMCHIMYSIDIFFNFDYKNLTFSLNIMESYLNEMTRFQLLN---QYQD 290
 L + I+ND +G+ S I + F + S+ +E++ F L+ Q

Sbjct: 182 LFRHINDECDVGIFDATNSTI-----ERRRIFIEVCSFNTDELSSFNLPILQVSC 235

289

Query: 291 IKISYTHYHFHDMFY-DYIKSFYRGGLNMYNTKYINKLIDPCFSID-INSSYPVVMYH 348
 S+ Y+ H+ +F DY+ Y + + + FS+D N + Y+ H
 Sbjct: 236 FNRSEIKYNIHNSFNEDYLDKPYELAIKDFAKRLKHYYSQFTPFSLDEFNQIHRYISQH 295

Query: 349 EKIPTWLYFYEHYSEPTLIPTFLDDNY 376
 E+I T L+F+ + + P L+ +Y
 Sbjct: 296 EEIDTSLFFFNVINAGVVEPHSLNQSHY 323

>gi|2258375|gb|AAD11909.1| (AF007261) transcription initiation
 factor sigma [Reclinomonas americana]
 Length = 532

Score = 39.9 bits (91), Expect = 0.070
 Identities = 49/205 (23%), Positives = 84/205 (40%), Gaps = 14/205 (6%)

Query: 100 NHFLKQDMRYFDNITRENIYLSAEENEHTLKMKEATILAKNQNVIL---EKRVKSSIN 156
 N+ + + P + ++IY+ + +KE L K NVI+ K +K N
 Sbjct: 177 NYLVKNSYLNLFKTVPHDSIYMNYSYIQTPLNKLKEYLQLIKIINVILQINKNIKKQN 236

Query: 157 LDLTMLNGFKFNIIDNFM---KTNTSIATLGKLLDGGYLTESQLKTDFTNYTIFDKND 213
 L+++FL F + N++ K + + + K L Y+T L T Y K
 Sbjct: 237 LNISLFLYKFYQELKWNVIFINKISRNTQKINIKLKNYSITFYNLITFIQYTTKKQRL 296

Query: 214 MNDSEAYDYAVKCFK--LTPEQLTYIHNDVILGMCHIHYSDFPNPDYN-KLTFSLNI 270
 D +K F K P+ +N+I G+ HI+ + N K+T I
 Sbjct: 297 KKDIFYKQIFIKTFLKQHKIPKINKIKNSLIKYGLTHIYDMILISILRENKIVTLKNRI 356

Query: 271 MESYLNEMTRFQLLNQYQDIKISY 295
 + +Y+ T + QY +KI Y
 Sbjct: 357 IFNYMPYITT---ISKQY--VKIGY 376

>gi|15734|emb|CAA37450| (X53370) DNA polymerase (AA 1-575)
 [Bacteriophage phi-29]
 Length = 575

Score = 39.5 bits (90), Expect = 0.092
 Identities = 41/150 (27%), Positives = 64/150 (42%), Gaps = 36/150 (24%)

Query: 497 LSKVVLNGLYG-----IPALRSHFNL-FRLDDNNELYNINIGYKXTERNIL--F 542
 L+K++LN LYG +P L+ + L FRL G + T+ +
 Sbjct: 381 LAKLMLNLSYKGFASNPDVTKVPYKENGALGRL-----GESETKDPVYTPM 429

Query: 543 STPVTSRSLYNLLVPFQYLTSEIDNFIYCDTOSLYMKSVMKPLNPSLFDPIALGKWD 602
 F+T+ + Y + Q D IYCDTDS+++ P + + DP LG W
 Sbjct: 430 GVFITAWARYTTITAAQACY-----DRIIYCDTDSIHLTGTEIPDVINKIDVPKILGYWA 484

Query: 603 IEENEQIDKMFVLNKKYAY----EVNGKI 627
 E+ ++ L K Y EV+GK+
 Sbjct: 485 HES-TFKRVKYLKQKTYIQDIYMKEVDGKL 513

Query= pt|110872 44AHJDORF002 Phage 44AHJD ORF |3789-5732|3 1
 (647 letters)

>gi|135273|sp|P27622|TAGC_BACSU TEICHOIC ACID BIOSYNTHESIS PROTEIN C
 >gi|478126|pir||D49757 techoic acid biosynthesis protein
 tagC - Bacillus subtilis (strain 168) >gi|143727
 (M57497) putative [Bacillus subtilis]
 >gi|2636103|emb|CAB15594.1| (Z99122) alternate gene
 name: dinC [Bacillus subtilis]
 Length = 442

Score = 112 bits (278), Expect = 7e-24
 Identities = 91/314 (28%), Positives = 147/314 (45%), Gaps = 58/314 (18%)

Query: 152 FELNLEPKFVMGFGGIRNAVNSINIDKETNHMYSTQSDS----QKPEGFWINKLTPSG 207
 F+ + PK V QS N D++ + +Y+TQ S + + I +L+ G
 Sbjct: 7 FDFNTNITPKLFTELRVADKTVLQSFNFDEKNHQIYTTQVASGLGKDNQSYRITRLSLEG 66

Query: 208 DLISSMRIVQGGHGTIGLERQSGEMKIWLHHD-----GVAKLLQVAYKDNVYLDLEA 262
 + SM + GGHGT IG+E + NG + IW +D ++L+ YK LD E +
 Sbjct: 67 LQLDSMLLKHHGGHGTNIGIENR-NGTIYIWSLYDKPNETDKSELVCFPPYKAGATLD-ENS 124

290

Query: 263 KGLTDYTPQSLLNKHTFTPLIDEANDKLILRFGDGTIQVRSRADVKNHIDNVEKEMTIDN 322
 K L ++ H TP +D N +L +R + D KN+ N ++ +TI N
 Sbjct: 125 KELQRFNSNMPF--DHRVTFALDMKNRQLAIR-----QYDTXKN--NNKQWVTIFN 170

Query: 323 SE-----NNDN-----RWMQGIADVGGDDLYWLSGNSSVNSHVQIGKYSLTGQKI 367
 + N +N ++QG +D LYW +G+++ S+ + +
 Sbjct: 171 LDDAIANKNNPLYTINIPDELHYLQGFFLDGGLYWTGDTNSKSYPNL-----ITV 222

Query: 368 YDYPPKLSYQDGINFPRD-----NFKEPEGICITYTNPKTKRKSLLAMTNGGGGKRFH 420
 +D K+ Q I +D NF+EPEGIC+YTNP+T KSL++ +T+G G R
 Sbjct: 223 FDSDNKIVLQKEITVGKDLSTRYENNFRPEGICMYTNPETGAKSLMVGITSGKEGNRIS 282

Query: 421 NLYGFFQLGEYEHF 434
 +Y + YE+
 Sbjct: 283 RIYAYH---SYENF 293

>gi|142847 (M64050) DNase inhibitor [Bacillus subtilis]
 Length = 125

Score = 51.9 bits (122), Expect = 1e-05
 Identities = 35/116 (30%), Positives = 55/116 (47%), Gaps = 10/116 (8%)

Query: 152 FELNELEPKFVMGFGGIRNAVNSINIDKETNHMYSTQSDS----QKPEGFWINKLTPSG 207
 F+ + PK V QS N D++ + +Y+TQ S + + I +L+ G
 Sbjct: 7 FDFNTITPKLFTLRLVADKTVLQSFNFDEKNHQIYTTQVASGLKDNTOQSYRITRLSLEG 66

Query: 208 DLISSMRIVQGGHGTITGLERQSNNGEMKIWLHHD----GVAKLLQVAYKDNVLD 258
 + SM + GGHGT IG+E + NG + IW +D ++L+ YK LD
 Sbjct: 67 LQLDSMLLKHGGHGTNIGMENR-NGTYIWSLYDKPNETDKSELVCFPYKAGATLD 121

>gi|4038407 (AF103943) factor C protein precursor (Streptomyces
 griseus)
 Length = 324

Score = 39.1 bits (89), Expect = 0.10
 Identities = 61/269 (22%), Positives = 102/269 (37%), Gaps = 33/269 (12%)

Query: 172 VNQSNIDKETNHMYSTQSDSQKPEG---FWINKLTPSGDLISSMRIVQGGHGTITGLER 228
 V QS D ++ Q S P+ I +L SG+ + M ++ GHG +IG +
 Sbjct: 66 VQQSFTFDIVNRRLFVAQLKSGSPDDSGDLCTQLDFSGNKLGHMYLLGFGHGVSIGAQ- 124

Query: 229 QSNEMKIWLHHDGVAKLLQVAYKDNVLDLEAKGLTDYTPQSLLNKHTFTF----- 281
 + +W D + + + + G T S L KH P
 Sbjct: 125 PVGADTYLWTEVD-----VNSNARGTRLARFKWNGATLSRTSSALAKHQPVPGATEMTC 179

Query: 282 LIDEANDKLILRFGDGTIQVRSRADVKNHIDNVEKEMTIDNSENNDNRWMQGIADVGGDL 341
 ID N+++ +R+ + + +V + V + D QG A+ G +
 Sbjct: 180 AIDPVNRMARILYLTASGRRYGIYNVADIAAGVYDKPLSDVPHPTGLTGFQGYALYGSYV. 239

Query: 342 YWLSGN-----SSVNSHVQIGKYSLTGQKIYDYPPKLSYQDGINFPRDNFKEPEGIC 394
 Y L+GN + NS+V + TG + + + G F+EPEG+
 Sbjct: 240 YQLTGNPYGPDNPNPGNSYVS--SVDVNTGALVQ----RAFTRAGSTL---TFREPEGMG 290

Query: 395 IYTNPKTKRKSLLAMTNGGGGKRFHLY 423
 IY + + L L +G G R NL+
 Sbjct: 291 IYRTAAGEVR-LFLGFASGVAGDRRSNLF 318

Query= pt|110873 44AHJDORF003 Phage 44AHJD ORF |6626-8389|2 1
 (587 letters)

>gi|138123|sp|P04331|VG9_BPPH2 TAIL PROTEIN (LATE PROTEIN GP9)
 >gi|75850|pir|WMBPT9 gene 9 protein - phage phi-29
 >gi|215327 (M14782) tail protein [Bacteriophage phi-29]
 >gi|225364|prf||1301270D gene 9 [Bacillus sp.]
 Length = 599

Score = 92.4 bits (226), Expect = 8e-18
 Identities = 126/618 (20%), Positives = 251/618 (40%), Gaps = 71/618 (11%)

Query: 5 TNFKFFYNTPFT-DYQNTIHFNSNKERDDYFLNGRHFKSLDYSKQPY-NFIRDMEINVD 62
 TN + + PF+ DY+NT F S+ + ++P R + + SK + F ++ ++V
 Sbjct: 9 TNVRILADVPSNDYKNTWFTSSSNQYNWF--NRKSRVYEMSKVTFMGFRENKPYVSVS 66

Query: 63 MQWHAQGINYMTFLS-DFEDRRYYAFVNQIEYVNDVVVKIYFVIDTIMTYTQGNVLEQL 121
 + +Y+ F + D+ ++ +YAFV ++E+ N V ++F ID + T+ ++
 Sbjct: 67 LPIDKLYSASYIMFQADYGNKWFYAFVTELEFKNSAVTYVHFEIDVLQTMFDMKQES 126

Query: 122 SNVNIERQHLSKRTYNYMLPMLRNDDVLKVSNNQVYVYNMQQYLENLVLQSSADLSKK 181
 I R+H+ K + P + D+ L ++ + + + +F S
 Sbjct: 127 F---IVREHV-KLWNDDGTPINTIDEGLSYGSEYDIVSVENHKPYDDMMFLVIISKSIM 182

Query: 182 FGT--KKEPNLDTSGTIYDNITSPVNLVMEYGDFFNFMKMSAYPWITQNFQK---V 235
 GT ++E L+ ++ + + P+ Y+ + + D +I N V
 Sbjct: 183 HGTSGEESRLNDINASL-NGMPQPLCYIHPF-----YKDGKVPKTYIGDNNANLSPIV 236

Query: 236 QMLPKDFINTKOLEDVKTSEKITGLKTLKQGGKSKWSLK-DLSL-----SFSNLQ 285
 ML P + D+ + +T LK K+ + LK D + N+
 Sbjct: 237 NMLTNIFSQSAVNDI-VNMYVTDYIGLKLDYKNGDKELKLDKDMFEQAGIADDKHGNVD 295

Query: 286 EMMLSK-----KDEFKHMIRNEYMTIEFYDWNNGNTMLLDAGKISQK 326
 + + K KD+ ++ Y E D+ GN M L I+
 Sbjct: 296 TIFVKKIPDYALEIDTGDKGWGFTHQESKLMMPYCVTEITDFKGNHMLKTEYINNS 355

Query: 327 TGVKLRKTSIIGYHNEVRVYPVDYNSAENDRPILAKNKEILIDTGSFLNTNITFNSPAQV 386
 +K++ + +G N+V DYN+ D + N+ S +N N
 Sbjct: 356 K-LKIQVRGSLGVSNNKVAYSVQDYNA---DSALSGGNRLTASLOSSLINNNFN----- 404

Query: 387 PILINNGILQSSQANRQ--KNAESQLITNRIDNVLNG---SDPKSRFYDAVSVASNLSP 441
 I I N L Q N+ +N +S ++ N I ++ G + + A+ +AS++
 Sbjct: 405 DIAILNDYLSAYLQGNKNSLENQKSSILFNGIMGMIGGGSAGASAAGGSALGMASV-- 462

Query: 442 TALFGKFNEEYNYFYKQQAQYKDLALQPPSVTESEMGNAFQIANSINGLTMKISVPSPK 501
 T + + QA+ D+A PP +T+ AF N G+ + +
 Sbjct: 463 TGMTSTAGNAVLQMQAMQAKQADIANIPPLTKMGGNTAFDYGNGYRGVYVIKQLKAEY 522

Query: 502 ITFLQKYMYLFGFEVDYNSFIPIINSMTVCNLYKCTGTITRIDPMLMEQLKALESG 561
 L ++ +G+++N + + NY++ + DI+ +++++ I ++G
 Sbjct: 523 RRLSSFFHKYGYKINRVKK--PNLRTRKAFNYVQTKDCFISGDINNNDLQEIRTIFDNG 580

Query: 562 VRFWHDGSGNPMQLQNPL 579
 + WH D GN ++N L
 Sbjct: 581 ITLWHTDNIGNYSVENEL 598

>gi|138124|sp|P07534|VG9_BPPZA TAIL PROTEIN (LATE PROTEIN GP9)
 >gi|75849|pir||WMBP92 gene 9 protein - phage PZA
 >gi|216058 (M11813) tail protein [Bacteriophage PZA]
 Length = 599

Score = 81.9 bits (199), Expect = 1e-14
 Identities = 127/618 (20%), Positives = 248/618 (39%), Gaps = 71/618 (11%)

Query: 5 TNKFPFYNTFPT-DYQNTIHFNSKERDDYFLNGRHFKSLDYSKQPYNFIRDME-INVD 62
 TN + + PF+ DY+NT F S+ + ++F + + SK + R+ I+V
 Sbjct: 9 TNVRILADVPPSNDYKNTRWPTSSSNQYNWF--NSKTRVYEMSKVTFGGFRENKSYISVS 66

Query: 63 MQWHAQGINYMTFLS-DFEDRRYYAFVNQIEYVNDVVVKIYFVIDTIMTYTQGNVLEQL 121
 ++ +Y+ F + D+ ++ +YAFV ++EY N ++F ID + T+ N+ Q
 Sbjct: 67 LRLDLLYNASYIMFQADYGNKWFYAFVTELEYKQVGTITYVHFEIDVLQTM-MFNKQES 125

Query: 122 SNVNIERQHLSKRTYNYMLPMLRNDDVLKVSNNQVYVYN--MQQYLENLVLQSSADLS 179
 S I R+H+ K + P + D+ L ++ + + + Y + + L S +
 Sbjct: 126 SF--IVREHV-KLWNDDGTPINTIDEGLNYGSEYDIVSVENHRPYDDMMFLVVISKIM 182

Query: 180 KKFGTKKEPNLDTSGTIYDNITSPVNLVMEY-----GD-----PINPMDK 221
 + E L+ ++ + + P+ Y+ + GD +N +
 Sbjct: 183 HGTAGEAESRLNDINASL-NGMPQPLCYIHPFYKDGKVPKTFIGDNNANLSPIVNMLTN 241

Query: 222 MSAYPWITQNFQKQMLPKDFINTK-----DLEDVKTSEKITGLKTLKQGGKSKWS 273
 + + N V M D+I K +L+ K + G+ KG +
 Sbjct: 242 IFSQKSAVNNI--VNMYVTDYIGLKLDYKNGDKELKLDKDMFEQAGIADDKHGNVDITFV 299

Query: 274 LKDL---SLSFNLQEMMLSKKDEFKHMIRNEYMTIEFYDWNNGNTMLLDAGKISQKTGVK 330
 K +L + KD+ ++ Y E D+ GN M L I +K
 Sbjct: 300 KIPDPVETLEIDTGKGGFTKQESKLMMPYCVTEVDFKGNHMLKTEYIDNNK-LK 358

Query: 331 LRTKSIIGYHNEVRVYPVDYNSAENDRPILAKNKEILIDTGSFLNTNITFNSPAQVPILI 390
 ++ + +G N+V DYN+ + L+ + L+T++ N+ + I+

292

Sbjct: 359 IQVRGSLGVSNKVAYSIIQDYNAGGS----LSGDRLTAS----LDTSLINNNPNDIAII- 409

Query: 391 NNGILGQSQQANRQ--KNAESQLITNRIDNVINGSDPKSRFYDAVSASNLSP----- 441
 N L Q N+ +N +S ++ N I +L G A + A SP

Sbjct: 410 -NDYLSAYLQGNKNLENQKSSILFNGIVGMLGG-----VSAGASAVGRSPFGLASSV 462

Query: 442 TALFGKFNEEYNYFYKQQQAEYKDLALQPPSVTESEMGNFQIANSINGLTMKISVPSPE 501
 T + + QA+ D+A PP +T+ AF N G+ + +

Sbjct: 463 TGMTSTAGNAVLEDMQALQAKQADIANIPPQLTKMGNTAFDYGNGYRGVYVIKQLKAEY 522

Query: 502 ITFLQKYMYLFGFEVNDYNSFIEPINSMTVCNYLKCTGTYTIRIDPMLMEQLKAILSEG 561
 L ++ +G+++N + + NY++ + DI+ +++++ I ++G

Sbjct: 523 RRSLSFFHKYGYKINRVKK--PNLRTRKAYNYIQTQKCFISGDINNNDLQEIRTIFDNG 580

Query: 562 VRFNHNDGSGNPMLQNPL 579
 + WH D GN ++N L

Sbjct: 581 ITLWHTDDIGNYSVENEL 598

>gi|1429238|emb|CAA67657| (X99260) tail protein [Bacteriophage B103]
 Length = 598

Score = 77.6 bits (188), Expect = 2e-13
 Identities = 130/623 (20%), Positives = 240/623 (37%), Gaps = 86/623 (13%)

Query: 5 TNFKFFYNTPT-DYQNTIHFNSKERDDYFLNGRHFPSLDYSKQPYNFI---RORMEIN 60
 T+ + F N PP+ DY++T F + + YF + K + NF+ I

Sbjct: 9 TDVRIFSNVFPFNDYKSTRWFTNADAQYSYF---NAKPRVHVINECNFVGLKEGTPHIR 64

Query: 61 VDMQWHDAGQINYMFLS-DFEDRRYYAFVNQIEYVNDVVVKIYFVIDTIMTYTQGNVLE 119
 V+ + D YM F + + ++ +Y FV ++EYVN V +YF ID I T+ +

Sbjct: 65 VNKRIDDLNACYMIFRNTQYSNKFYCFVTRLEYVNSGVNLYFIDVIQTN-MEDFKF 123

Query: 120 QLSNVNIERQHLSKRTYNYMLPMLRNDDVLKVSNNQYVYNQMQYLENLVLFQSSADLS 179
 Q S + E Q + P+ D+ L + V Q ++F S

Sbjct: 124 QPSYIVREHQEMWDANNE---PLTNTIDGLNYGTEDVVAVEQYKPYGOLMFMVCISKS 180

Query: 180 KKGFTKKEPNLDTSGKTIYDNITS---FVNLYVMEYGDFFINPMDKMSAYFWITONFQKVQ 236
 K T E G I N I P++ YV + + D S P +T +VQ

Sbjct: 181 KMHTATAGET---FKAGEIAANINGAPQFLSYVHPF-----YEDGSS--PKVTIGSNEVQ 230

Query: 237 ML-PKDFINTKLEDVKTSEKITGLKT-----LKQGGKSKENSLKDLSPSNL----- 284
 + P DP+ ++ + ++ T + +K SL+D + +

Sbjct: 231 VSKPTDFLKNMPTQEHAVNNIVSLYVTDYIGLNIHYDESAKTMSLRDTMFEHAQIADDKH 290

Query: 285 -----QEMMLSKKDEPKHMIRNEYMTIEFY-----DWMGNTMLLDAGK 322
 +E + +F NE + Y D+ GN + +

Sbjct: 291 PNVNTIYLKEVKEETIDTGYKFASFANNEQSKLLMYPYCVTTITDFKGNQIDIKNEY 350

Query: 323 ISQKTGVKLRTKSIIGYHNEVRVYPVDYNS---AENDRPILAKNKEILIDTGSFLNTNIT 379
 ++ + +K++ + +G N+V DYN+ D+ + A NT++

Sbjct: 351 VNG-SNLKIQVRGSLGVSNKVITYSVQDYNADTTLSGDQNLTA-----CNTSLI 398

Query: 380 FNSPAQVPILINNGILGQSQQANRQ--KNAESQLITNRIDNVLN---GSDPKSRFYDAVS 434
 N+ V I+ N L Q N+ +N + ++ N + ++L G+ + AV

Sbjct: 399 NNNPNDVAII--NDYLSAYLQGNKNLENQKDSILFNGVMSMLGNGIGAVGSAATGSAVG 456

Query: 435 VASNLSPALFGKFNEEYNYFYKQQQAEYKDLALQPPSVTESEMGNFQIANSINGLTMKI 494
 VAS S T + + QA+ D+A PP + + A+ N G+ +

Sbjct: 457 VAS--SATGMVSSAGNAVLIQGMQAKQADIANTPPQLVQMGNTAYDYGNGYRGVYVIK 514

Query: 495 SVPSPEKITFLQKYMYLFGFEVNDYNSFIEPINSMTVCNYLKCTGTYTIRIDPMLMEQL 554
 + L + +G++ N + + + NY++ I +++ +

Sbjct: 515 KQIKEZYRNILSDFSRKYGYKTNLVK--MPNLRRESYNYVQTKDCNIIGNLNEDLQKI 572

Query: 555 KAILSGVRFVHNDGSGNPMLQN 577
 + I +SG+ WH D G+ L N

Sbjct: 573 RTIFDSGITLWADPVGDYTLNN 595

>gi|215339 (M12456) p9 tail protein [Bacteriophage phi-29]
 >gi|224163|prf||1011232C protein p9,tail [Bacteriophage
 phi-29]
 Length = 335

293

Score = 71.0 bits (171), Expect = 2e-11
Identities = 64/293 (21%), Positives = 123/293 (41%), Gaps = 20/293 (6%)

Query: 292 KDEFKHMIRNEYMTIEFYDWNNGNTMLLDAGKISQKTGVLRTKSIIGYHNEVRVYPVDYN 351
KD+ ++ Y E D+ GN M L I+ +K++ + +G N+V DYN
Sbjct: 57 KDQESKLMYPYCVTEITDFKGNHMLKTEYINNSK-LKIQVRGSLGVS NKVAYSVDYN 115

Query: 352 SAENDRPILAKNKEILDTGSLNTNITFNSFAQVPILINNGILGQSQQANRQ--KNAES 409
+ D + N+ S +N N I I N L Q N+ +N +S
Sbjct: 116 A---DSALSGGNRLTASLDSSLINNNPN-----DIAILNDYLSAYLQGNKNSLENQKS 165

Query: 410 QLITNRIDNVLNG---SDPKSRFYDAVSANLSPTALFGKFNEEYNFYKQQQAEYKDLA 466
++ N I ++ G + + A+ +AS++ T + + QA+ D+A
Sbjct: 166 SILFNGIMGMIGGGISAGASAAGGSALGNASSV--TGMTSTAGNAVLQMAMQAKQADIA 223

Query: 467 LQPPSVTESEMGNAPQIANSINGLTMKISVSPKEITFLQKYMLFGFVNDYNSFIEPI 526
PP +T+ AF N G+ + + L ++ +G+++N +
Sbjct: 224 NIPPQLTKMGNTAFDYNGYRGVYVIKKQLKAEYRSLSSFPFKYGYKINRVKK--PNL 281

Query: 527 NSMTVCNYLKCTGTYTIRIDPMLMEQLKAILESGVRFWHNDGSGNPMLQNPL 579
+ NY++ + DI+ +++++ I ++G+ WH D GN ++N L
Sbjct: 282 RTRKAFNYVQTKDCPISGDINNDLQEIIRTIFDNGITLWHTDNIGNYSVENEL 334

>gi|1181968|emb|CAA87738.1| (Z47794) tail protein [Bacteriophage
CP-1]
Length = 230

Score = 53.9 bits (127), Expect = 3e-06
Identities = 29/113 (25%), Positives = 54/113 (47%), Gaps = 3/113 (2%)

Query: 1 MRKLNTFKFFYNTFF-TDYQNTIHFNSNKERDDYFLNGRHPKSLDYSKQPYNFIRDRMEI 59
M++ T + +PF DY N I+P + + +D+P + Y + + + I
Sbjct: 1 MQESTKIWLKSPFKNDYANVINFETRESMEDDFTKKNPHIEIVYEYDKPYQTQRNGSI 60

Query: 60 NVDMQWHDAAQGINYMTFLSDFEDRRYYAFVNGIEYVNDVVKIYFVIDTIMTY 112
V + + + YM P+++ R YYAFV + Y+N+ +I + +D TY
Sbjct: 61 VVSGRVEKYENVTYMRFINN--GRYYAFVFDVLYINEDATRIIYEVVDWNTY 111

>gi|1181970|emb|CAA87740.1| (Z47794) tail protein [Bacteriophage
CP-1]
Length = 586

Score = 42.2 bits (97), Expect = 0.010
Identities = 79/381 (20%), Positives = 139/381 (35%), Gaps = 92/381 (24%)

Query: 277 LSLSPSNLQEMMLSK--KDEPK---HMIRNEYMTIEFYDWNNGNTMLLDAG---KISQKT 327
L +++ +QE + S KD+ + ++ +E+ IE YD GN+ + I +
Sbjct: 187 LKIAVDQIQEGLRSGYMKDDLEIEVQLNSEFTETELYDIYGNVYVYQPYLPRTIDEAH 246

Query: 328 GVKLRTKSIIGYHNEVRVYPVDYNSAEN----DRPIL----- 360
K+ +G N+V + ++YN+A N D+ IL
Sbjct: 247 KYKVIVSGSLGDSNQVHINFLEYNNAMNVSYADKNILDSLESGDWAENHPEHFKYGLNDV 306

Query: 361 -AKNKEILDT-GSFLNTNITFNSFAQVPILINNGILGQSQQANRQKNAESQLITNRIDN 418
K+ IL D S++ ++ Q+ N +L QS + + + A + +
Sbjct: 307 TGKSVAILNDAAEASYIQSHKNQMEHTQLTFKENRDMKQSVDSLKNQVATANSQASYNQ 366

Query: 419 VLNGSDPKSRFYDAVSANLSPTALFGKF-----NEEYNFYKQQQ-- 459
S + + + + S N++ L G F N +YN QQ
Sbjct: 367 FAVDSANINQWTEGASGILNVAGNLLTGNFGGALGLASGGMKVFMANRDYNDKVVOQGF 426

Query: 460 -----AEYKDLALQPPSVTESEMGNAPQIANSIN 488
A DL QP SV + AFQ N +
Sbjct: 427 TSENNALKSQSNALANMKSKIALDQSIKAYNATMADLQNPISVQIQGNDLAFQSGNRLT 486

Query: 489 GLTMKISVSPKEITFLQKYMLFGFVNDY-NSFIEPINSMTVCNYLKCTGTY--TIRD 545
+ K+S+ + + +Y +G VN + N + + S NY+K T+R
Sbjct: 487 DVYKVKSLAQKEIMGRANEYIKCYGVLVNWFTNDALSVMRSRKRFPNYIKMINVNLGTLR- 545

Query: 546 IDPMLMEQLKAILESGVRFWH 566
+ M ++AI +SGVR W+
Sbjct: 546 ANQSHMNAIQAIQFQSGVRIWN 566

Score = 52.3 bits (123), Expect = 6e-06
Identities = 59/246 (23%), Positives = 105/246 (41%), Gaps = 27/246 (10%)

Query: 407 KYLMKQ 412
+++K+
Sbjct: 1074 FFIKK 1079

Score = 49.2 bits (115), Expect = 5e-05
Identities = 67/287 (23%), Positives = 110/287 (37%), Gaps = 60/287 (20%)

Query: 359 YVVDNDRYLYLDMNKIIKPHIKNEMKQNMSEFERKEK-ITYEDNYIEN 404
V+N + ++IK + + N E+ + EK +Y + EN
Sbjct: 3808 ISVNTLNLCLNIIKELIKLNNNKKKIILNYYEYHKVEKLLYYRHSFEN 3854

Score = 35.6 bits (80), Expect = 0.70
Identities = 62/290 (21%), Positives = 121/290 (41%), Gaps = 65/290 (22%)

Query: 120 QDKEIGVITDLNSATDLKYHSNFLKHYPIIIYDEFL-----ALEDDYLIIEWDKLKIYE 174
+ EI ITD++ +YH N+LK + +E++ + +D + DE ++T+ E
Sbjct: 4524 KQNEINNITVDYGNKKEYHENYLKVKQNKVNEEYIETFKSDKDCSIKDEACTIRTLSE 4583

295

Query: 175 S--IDRNHGNVDYIGFPMFLGNVNFSSPILSNLNIYNLLQKHOM--TSRLYKNIFL 230
 S I N N+D + + + S P N++ N++K+ +N R+ KN
 Sbjct: 4584 SCNISENISNID-----MDDEDHISFPNGRNVHDNNYMKKNHVNYDKMRVGVKNKIP 4634

Query: 231 EMRRNDYVNEKRNTAFNSNDDAMTGEFNEYNLADDNLRNHINQNGD 280
 D + +++ + +D M++ ++ E ++ + L + NG+
 Sbjct: 4635 SFTHFDKILDEKKKK----SKDMSSSKWLEREHIKEIKLEKNEYMNGN 4680

Score = 34.0 bits (76), Expect = 2.0
 Identities = 47/211 (22%), Positives = 84/211 (39%), Gaps = 32/211 (15%)

Query: 210 IYNLLQKHOMTSRLYKNIFLEMRNDYVNEKRNTAFNSNDDAMTGEFNEYNLADD 269
 I++LLQK LY+N+ + R + N+ T E ++ + +
 Sbjct: 918 IFSLLQKSSPLLVLVYENVHI-----REGEKYGRNE--ATDNEVDYKKGDIKH 964

Query: 270 NLRNHINQNGDFFYIKTD---DKYIKVMYVTTFTMTNIIIVPYTKQYEFCTKIRDIDNHV 326
 N+ N + D + D+ K MY + V E K D+ N+
 Sbjct: 965 NVTNEHGNHSDSYPGNSLNRKPKMYE-DIYKEKGFVKSDCSNIEI--KKNDMINND 1021

Query: 327 TYLRDDMFYKENMERYYYNPSNLHFDNAYSKNYVVDNDRYLYLDMNKII----KPHIKNE 382
 Y +++ FY+++ Y+ + YV++ +YL +N ++ F +KN+
 Sbjct: 1022 VYKKN--PYEDSRINMIYDEDEIKTWFLIPKHYVIN---IIYFLNILLTDESNFKLKNK 1077

Query: 383 MKQNMSEFERKEKIYEDN-----YIENTKKY 408
 E K IYEDN +N KKY
 Sbjct: 1078 KYGYFVNBETKGTIYEDNGLQEILKNGKKY 1108

Score = 33.6 bits (75), Expect = 2.7
 Identities = 42/198 (21%), Positives = 77/198 (38%), Gaps = 42/198 (21%)

Query: 222 SRLYKNIFLEMR---RNDYVNEKRNTAF-----NSNDDAMTGEFNEYNLA 267
 S LY I++ + +N + K+NT + N+++D TT E + +
 Sbjct: 411 SVLYSIYMNKKYKKNPIITNKQNTNVYPENDVIQLSVENTSSEDFTTNTRESSLSGM 470

Query: 268 DDNLRNHINQNGDFFYIKTDKDYIKVMYVTTFTMTNIIIVPYTKQYEFCTKIRDIDNHVT 327
 +++R +N D +DDK ++Y N YTK E
 Sbjct: 471 MNDMRYSVNNYADEKVYHSDDKSDHLIYKHVHDEKNKYDEMYTKTKE----- 517

Query: 328 YLRDDMFYKENMERYYYNPSNLHFDNAYSKNYVVDNDRYLYLDMNKIIKPHIKNEKQNM 387
 +++ YK N+ + N K LD+ K I H+KN+ + N
 Sbjct: 518 --NENIYKSNIVDKKTCDISSEMVGKDK-----LDVEKYIGSHVQND--ENNK 563

Query: 388 SEFERK-EKIYEDNYIEN 404
 + ++K + + + YI+N
 Sbjct: 564 EKLKKKIDNVNKKYIDN 581

>gi|3845297 (A8001421) hypothetical protein [Plasmodium falciparum]
 Length = 2380

Score = 48.0 bits (112), Expect = 1e-04
 Identities = 87/390 (22%), Positives = 160/390 (40%), Gaps = 65/390 (16%)

Query: 20 VRKKIPDKYNQIELVDELMDDDIDYISISNRSDGKSPNYVSFF-----IYLAKLDIKF 74
 +++K +K ++ + +N D + ++ R K+ NY++ +YL I DI
 Sbjct: 1049 LQRKNMKNKSKNRNRNRYINKDSNIHLMNLIRIKFKNLNYMMNSFEIELYLKINNDIFL 1108

Query: 75 TLLSRHYTLRDAYR-----DFIEEIIDEN-PLFKSKRVTFRSARDYLAIIYQDKEIGVI 127
 +Y +++ Y + + + EN + +++ ++ + Y +K+
 Sbjct: 1109 QFNKHNYNVQNFYNSITLINIMSKYYSNFYAYNLEKIVYKFLNNKNPEYIEKQYSSK 1168

Query: 128 TDLSATDLKYHSNFKLHYPIIIYDEFLA---LEDDYLIDWDKLKTIYESIDRNHGNV 183
 D+N D+ ++ +K+ II EFL L+ D I + KLKT ++
 Sbjct: 1169 EDMNEL-DILVNTYDMKYDKII---EFLKNGYKIDRYIYFYPKLKT-----DI 1214

Query: 184 DYIGFPMFLGNVNFSSPILSNLNIYNLLQKHOMTSRLY-----KNIF--LEMRRN 235
 F ++FL N+ L NI +++ K + Y K IF + M+ +
 Sbjct: 1215 ILFFFKEIFLNDNLIKIDRKFLKK-NITIMIEVLKEIFFKEYVKRCITKVIFFPVMKEH 1273

Query: 236 DYVNEKR-----NTRAFNSNDDAMTGEFNEYNLADDNLRNHINQNGDFFYIKTD 287
 D+V K N+ FN+ D + N YN D+ N+ N N +Y K

296

Sbjct: 1274 DHVMNKQYNNQYVNNNSNMFNTRGDHNNMNQTNNDHYNHHYDDTHNNNNNNNSKYK-KNK 1332
 Query: 288 DKYIKVMYNTTFTMTNIIV---VPYTKQYEFCTKIRDIDNHVTYLRDDMFYKEN----ME 340
 +K K+MY +++ + V K + K I + Y+ ++ N +
 Sbjct: 1333 NQN-KIMYEKERSSSLPISNNVQDVKPIKHYLYSSIIYKNFYIYIIEIKNFNNKITKIN 1391
 Query: 341 RY-YYNPSNLHFDNAYSQNYVVDNDRYLYL 369
 RY YYN NL+ D+ ND YL+L
 Sbjct: 1392 RYNYNYMNLNIDDL-----NDAYLFL 1413

Score = 32.5 bits (72), Expect = 6.0
 Identities = 46/183 (25%), Positives = 73/183 (39%), Gaps = 26/183 (14%)

Query: 225 YKNIFLEMRNDYVNEKRNTRAFNSNDDAMTGEFEFNEYNLADDNLNRHINQNGDFFYI 284
 +KNI ++ ++N + NSN + + N N+ +N N IN + I
 Sbjct: 27 HKNINKIKKKFIPINDSNMNCNNSNSNSNNNNNNNNNIVRNN--NNFINADKKKQVI 85
 Query: 285 KTDDKYIKVMYNTTFTMTNIIVVPYTKQYEFCTKIRDIDNHVTYLRDDMFYKENMERYYY 344
 +D IK V NI Y ++ + D+ N+ + + KE ER
 Sbjct: 86 LNEDDDIKNKELVDESFFVNIFF--YENYFKNLFLNDVSNKVI--NIEQKEGDER--- 138
 Query: 345 NPSNLHFDNAYSQNYVVDNDRYLYLDMNKIKFIKHEMKNQMSSEFERKEKIYEDNYIEN 404
 N N N +KN V DN +NK IKV +N++E Y N++ +
 Sbjct: 139 NADN---NLKKNIVRDN-----INK-----IKN--TRNVNEILIYNNKYIINFLND 180
 Query: 405 TKK 407
 T K
 Sbjct: 181 TTK 183

>gi|4493936|emb|CAB38972.1| (AL034556) predicted using hexExon;
 MAL3P5.6 (PFC0600w), Hypothetical protein, len: 250 aa
 [Plasmodium falciparum]
 Length = 249

Score = 47.3 bits (110), Expect = 2e-04
 Identities = 53/215 (24%), Positives = 87/215 (39%), Gaps = 30/215 (13%)

Query: 209 NIYNLLQKHKMTSRLYKNIIFLEMRNDYVNEKRNTRAFNSNDDAMTGEFEF--NEYNL 266
 NIYN L++ YKN N ++ +N N+N EFE N YN
 Sbjct: 13 NIYNKLEEK-----YQVFLKLQNMNSHMGASQNMV--NNYTMNELEEFKINNNYNN 64
 Query: 267 ADDNLNRHINQNGDFFYIKTD----DKYIKVMYNTTFTMTNIIVVPYTKQYEFCTKIRD 321
 ++N+ N+IN D+ IK +K ++ YN + I T +++
 Sbjct: 65 NNNNNNNNNYDYDMNIKVSQSVQHNRQLQDFYNNKNSFQHYIKLKTCTCPDADDIRNL 124
 Query: 322 IDNHVTYLRDDMFYK-----ENMERYYYNPSNLHFDNAYSQNYVVDNDRYLYLDMNKIK 376
 ++ + Y RD+ K EN + N + N+ S NY DN+ LY +N++ K
 Sbjct: 125 LEKRLAYERDMTLIKNIQEBENKKGIGINGNFGSESNSSSSNY--DNMYLLYRKINRLNK 182
 Query: 377 PHIKNEMKNMSEFERKEKIYEDNYIENTKKYLMK 411
 + ++ KI KKY++K
 Sbjct: 183 TNTNKSQNRSRKRRINSKI-----DKKYIK 209

>gi|3845165 (AE001390) hypothetical protein [Plasmodium falciparum]
 Length = 1247

Score = 45.7 bits (106), Expect = 6e-04
 Identities = 52/239 (21%), Positives = 94/239 (38%), Gaps = 38/239 (15%)

Query: 206 SNLNIYNLLQKHKMTSRLYKNIIFLEMRNDYVNEKRNTRAFNSNDDAMTGEFEFNEYN 265
 +N N +N ++K K R I +N + +N ++N+D E N N
 Sbjct: 474 NNTNKNWEIKKKKKFKRBRVKIINNSFQNEAEDDKNNNNNDNNNDNNDNNNNNNNEN 533
 Query: 266 LADDNLNRHINQNGDFFYI-KTDDKYIK---VMYNTTFTMTNIIVVPYTKQYEFCTKIR 320
 D+N N+ + N D I D+ Y +YN T ++ YTK + + +
 Sbjct: 534 NNDNNNNNDINNDINNDINNDNNYNNNDINNLNEMTKKCMNDNSYTKYFFYIFTL- 592
 Query: 321 DIDNHVTYLRDDMFYKENMS-----RYYN-----PSNLHFDNAYS 356
 + + ++ + FY++N + ++YN + N
 Sbjct: 593 ---DMLPSIKFETFYEKNTDHNFNENYKFFYNTDDTDIINAIKQNVKNKKQGNIVI 649
 Query: 357 KNYVVDNDRYLYLDMNKIKFIKHEMKNMSEFER---KEKIYEDNYIENTKKYLMK 411
 KNY+ N+ Y YL+ N+ + I + K +E K+ I+ ++Y E K K

Sbjct: 650 KNYINHNE-YSYLEYNENKNYEINKKEKLLTENYEYDMYIKDNIHYNDYSEGDGKQTKK 707

Query: 207 NLNIYNLLQKHKMTSRLYKNIFFLEMRNDYVNEKRNTRAFNSDDAMTTGEFEFNEYNL 266
N+N+YN + K K Y F + D + + + N D E YN
Sbjct: 564 NINLYNEMTKKKCMLDNSYTKYFFYIFTLMLPSIKFETPYEKNTDHKNFNENYKFFYNT 623

Query: 267 ADD-----NLRNHINQNGDFF---YIKTDDK YIKVMYNVT-TFMTNIIIVVPYTKQ 312
 DD N++N +NG+ YI ++ Y + YN + N T+
 Sbjct: 624 DDDTDIINAIAKKNVKNK-KKGNIVIKFYINHNE-YSYLEYNENKNYEINKKEKLLTEN 681

Query: 313 YEFCTKIRDIDNHVTYLRDDMFYKENMERYYYNPSNLHFDNAYSK-----NYV--VD 362
YE+ I+D ++ Y D + + YN +N +N Y K +Y+ VD
Sbjct: 682 YEYDMYIKDNIHYNDYSEGDKGKTKASSFLYNNN---NNKYKKEDNKTQIISYMDHVD 738

Query: 363 NDR-----YLYLDMNKIIKFHIK-NEM---KKNMSEFERKEKIYEDNYIENTKKY 408
N+ Y + +++ F +K N+M K+ P +E I + +EN K+
Sbjct: 739 NENGVGKGLKKRNLFFYNNSDQLYNFDVKDNDMIKYEKRSQKNFVEEEFINGNRKMENEDKH 798

Query: 409 LMKQY 413
L K Y
Subject: 799 LKKHY 803

Query= pt|110877 44AHJDORF007 Phage 44AHJD ORF |2044-3027|1 1
(327 letters)

```
>gi|1181960|emb|CAA87731.1| (Z47794) connector protein
      [Bacteriophage CP-1]
      Length = 337
```

Score = 45.7 bits (106), Expect = 5e-04
Identities = 44/184 (23%), Positives = 84/184 (44%), Gaps = 13/184 (7%)

Query: 127 QIHKLYDNCMSGNFVVMQNKPIQYNSDIBIEHYTDELAEVALSRFSLIMQAKFSK--IF 184
 ++HK ++ +V+ N Y I +E++LA++L+ L A++ IF
 Shift: 125 ELHKDNPDKIKRCPVIPPNNNF-YEPYIGYLELFCFEKLADIET-IOLNRNAOITPYFIP 182

Query: 185 KSEINDESINQLVSEIYNGAPFVKMSPMFNAD-----DDIIDLTSNSVIPALTEMKR 236
N S+ + ++I N P V ++ + D D I + L ++

Subject: 183 ADNTNVL SMKNIENKIANFE PVVYL NKOKDODGDSFKOLSDYIOVFRTPADAPFLLDKLHD 242

Query: 237 EYQNKISELSNYLGINSLAVDKESGVSDDEEAKSNRGFTTSNSNIYLGREP-ITPLSKRY 295
E +++L ++GIN+ DK+ + EA SN G ++N + K R + ++K Y
Sbjct: 243 EKLRLVNMOLLTFIGINANPSDEKERLVVSEALSNNGVISANIEVGWKSRRKRFVELINKCY 302

Query: 296 GLDI 299
GL+I
Subject: 303 GLEI 306

```
>gi|1429239|emb|CAA67658| (X99260) upper collar protein
      (Bacteriophage 8103)
      Length = 308
```

Score = 44.9 bits (104), Expect = 8e-04
Identities = 40/159 (25%), Positives = 73/159 (45%), Gaps = 11/159 (6%)

Query: 150 YNSDIEI-----IEHYTDELAEVA-LSRFSLIMQAKFSKIFKSEINDESINQLVSEIYNG 203
 YN+D++ +E + +LAE+ + + Q I ++ N S+ + ++
 Sbict: 121 YNNDLKCSLTPALEMFAODLAEKLEIIAVNONAOKTPVLIAANDNNQLSLKNIYNOYEGN 180

Query: 204 APFVKMSPMFNADD-DIIDLTNSVIPALTEMKREYQNKISELSNYLGINS LAVDKESGV 262
 AP + + + D+ + + V+ L K N E+ YLGI + ++K+ +
 Subject: 181 APVIFVHESLDLNLKVPKTDAPYVVVDKLNQAKVAVN---EVMYLGIKIANLEKKERM 237

Query: 263 SDEEAKSNRGFTTSNSNIYLKGR-EPITFLSKRYGLDIK 300
E SN S+ NIYLK R E +S+ YGL++K
Sbjct: 238 VTSEVDSNDEOIESGNIYLKAROEACNKISELYGLNLK 276

>gi|137915|sp|P07535|VG10_BPPZA UPPER COLLAR PROTEIN (CONNECTOR PROTEIN) (LATE PROTEIN GP10) >gi|75851|pir|WMBP10 gene

298

10 protein - phage PZA >gi|216059 (M11813) upper collar
protein [Bacteriophage PZA]
Length = 309

Score = 43.8 bits (101), Expect = 0.002
Identities = 38/160 (23%), Positives = 75/160 (46%), Gaps = 13/160 (8%)

Query: 150 YNSDIEI-----IEHYTDELAELVALSRFLIMQAKFSKIF--KSEINDESINQLVSEIYN 202
YN+D+ +E + ELAE+ S+ A+ + + ++ N S+ Q+ ++
Sbjct: 122 YNNDMSFPTTPTLELFAELAEELK-EIISVNQNAQKTPVLIRANDNNQLSLKQVYNQYEG 180

Query: 203 GAPFVKMSPMFNADD-DIIDLTNSVIPALTEMKREYQNKISELSNYLGINS LAVDKESG 261
AP + ++D ++ + V+ L K N E+ +LGI + ++K+
Sbjct: 181 NAPVIFAHEALDSDSIEVFKTDAPYVVDKLNQKNAVWN---EMMTFLGIKNANLEKKER 237

Query: 262 VSDEEAKSNRGFTTSNSNIYLKGR-EPITFLSKRYGLDIK 300
+ +E SN S+ ++LK R E +++ YGLD+K
Sbjct: 238 MVTDEVSSNDEQIESSGTVFLKSREEACEKINELYGLDVK 277

>gi|137914|sp|P04332|VG10_BPPH2 UPPER COLLAR PROTEIN (CONNECTOR
PROTEIN) (LATE PROTEIN GP10) >gi|75852|pir||WMBPC9 gene
10 protein - phage phi-29 >gi|215328 (M14782) upper
collar protein (Bacteriophage phi-29) >gi|215340
(M12456) p10 connector protein (Bacteriophage phi-29)
>gi|224161|prf||1011232A protein p10,connector
[Bacteriophage phi-29] >gi|225365|prf||1301270E gene 10
[Bacteriophage phi-29]
Length = 309

Score = 41.4 bits (95), Expect = 0.009
Identities = 37/160 (23%), Positives = 75/160 (46%), Gaps = 13/160 (8%)

Query: 150 YNSDIEI-----IEHYTDELAELVALSRFLIMQAKFSKIF--KSEINDESINQLVSEIYN 202
YN+D+ +E + ELAE+ S+ A+ + + ++ N S+ Q+ ++
Sbjct: 122 YNNDMAFPTTPTLELFAELAEELK-EIISVNQNAQKTPVLIRANDNNQLSLKQVYNQYEG 180

Query: 203 GAPFVKMSPMFNADD-DIIDLTNSVIPALTEMKREYQNKISELSNYLGINS LAVDKESG 261
AP + ++D ++ + V+ L K N E+ +LGI + ++K+
Sbjct: 181 NAPVIFAHEALDSDSIEVFKTDAPYVVDKLNQKNAVWN---EMMTFLGIKNANLEKKER 237

Query: 262 VSDEEAKSNRGFTTSNSNIYLKGR-EPITFLSKRYGLDIK 300
+ +E SN S+ ++LK R E +++ YGL+K
Sbjct: 238 MVTDEVSSNDEQIESSGTVFLKSREEACEKINELYGLNVK 277

Query= pt|110878 44AHJDORF008 Phage 44AHJD ORF |3020-3775|2 1
(251 letters)

>gi|4982468|gb|AAD30963.2| (AF118151) SNF1/AMP-activated kinase
[Dictyostelium discoideum]
Length = 718

Score = 52.3 bits (123), Expect = 3e-06
Identities = 28/118 (23%), Positives = 56/118 (46%), Gaps = 5/118 (4%)

Query: 121 YLQSQGFTEHNEDTTSNTDETSQNAATSLDNSTGMTANRNAYV---SLPQSEVNIDVDN 176
+ + GF N ++ SN + +N N + N+ T N N + ++ + +N + +N
Sbjct: 382 FTTTGTGPNFTNSNSISNNNNNNNNNNNTTNNNNNTTNNNNSIINNINNINNINNINNINN 441

Query: 177 TTLRFADNNTIDNGKTVNKSSNESNQAKRNQKGNAGTQFTKQYLID-NIDKAYD 233
+NN I+N N ++N +N N N N N+ + T+ + I N++ +Y+
Sbjct: 442 NNNNNNNNNIINNINNINNINNINNINNINNINNINNINNSSISGGTEVFSISPNNNSYN 499

Score = 37.5 bits (85), Expect = 0.094
Identities = 17/111 (15%), Positives = 45/111 (40%)

Query: 130 HNEDTTSNTDETSQNAATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTIDN 189
+N + +N + +N N + +N++ ++ + P + + + + + N+ ++
Sbjct: 456 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSSISGGTEVFSISPNNNSYNSSNGSNGSNNNS 515

Query: 190 GKTIVNKSSNESNQAKRNQKGNAGTQFTKQYLIDNIDKAYDLRKKILN 240
N +N +N N N N N N ID+++ + + + N
Sbjct: 516 NNNNTNDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCIDSVNNSLNNENDVNN 566

Score = 32.8 bits (73), Expect = 2.4
Identities = 31/140 (22%), Positives = 57/140 (40%), Gaps = 14/140 (10%)

Query: 109 LNVVYSSEVEKYLSQGFTEHNEDTTS---NTDTSNQATSLDNSTGMTANRNAYVSL 165
LN Y+S+ S N+T+ N+ +N N+ +N+ N N+
Sbjct: 494 LNNYSNSSSGNSGNSNSNNNTNDNNNNNNNNNNNNNNNNNNNNNNNCIDS 553

Query: 166 PQSEVN--IDVDNTTLRFADNNTIDNGKTVNKSS-----NESQNAKRQKQGN 215
+ +N DV+N+ + +NN D+G N ++ N N+ N GN
Sbjct: 554 VNNSLNNENDVNNSNINNNNNNSDDGSNNNSYEGGGDVL LLLSDLNNGNQLGGNDNGNVV 613

Query: 216 GTQFTKQYLIDNIDKAYDLR 235
Q L++++D D++
Sbjct: 614 NLNNNFQ-LLNSLDLNSDIQ 632

Score = 31.7 bits (70), Expect = 5.4
Identities = 25/115 (21%), Positives = 48/115 (41%), Gaps = 10/115 (8%)

Query: 130 HNEDTTSNTDTSNQATSLDNST---GMTAN-RNAYVSLPQSEVNIDVDNTTLRFADNN 185
+N+ +N+ +N N+S+ T ++ N N+Y S S N+ N+ +N
Sbjct: 462 NNNNNNNNNNNNNNNNNSSISGGTEVFSISPNNNSYNS--NSSGNSGNSNSNNNNNT 519

Query: 186 TIDNGKTVNKSSNESQNAKRQKQGNAGTQFTKQYLIDNIDKAYDLRKKILN 240
DN N ++N +N N N N N + ++++ D+ +N
Sbjct: 520 NNDN---NNNNNNNNNNNNNNNNNNNNNNNCIDSVNNSLNNENDVNNSNIN 570

Score = 31.7 bits (70), Expect = 5.4
Identities = 15/104 (14%), Positives = 43/104 (40%)

Query: 110 NVVYSSEVEKYLSQGFTEHNEDTTSNTDTSNQATSLDNSTGMTANRNAYVSLPQSE 169
N+ +++ + + +N+ +N+ +N N+ +N+ + + V
Sbjct: 434 NNNSSISGGTEVFSISP 493

Query: 170 VNIDVDNTTLRFADNNTIDNGKTVNKSSNESQNAKRQKQGN 213
+N ++ + ++ +N N +++ +N N N N N
Sbjct: 494 LNNYSNSSSGNSGNSNSNNNTNDNNNNNNNNNNNNNNNNNN 537

Score = 30.9 bits (68), Expect = 9.2
Identities = 16/84 (19%), Positives = 34/84 (40%)

Query: 130 HNEDTTSNTDTSNQATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTIDN 189
+N+ +N+ +N N+ +N+ + S+ + N N++ +N+ +N
Sbjct: 455 NNNSSISGGTEVFSISPNNNSYNSSSGNSGNSNN 514

Query: 190 GKTVMKSSNESQNAKRQKQGN 213
+ N +N N N N N
Sbjct: 515 SNNNTNDNNNNNNNNNNNNNNNNNNNN 538

>gi|1730077|sp|P18160|KYK1_DICDI NON-RECEPTOR TYROSINE KINASE SPORE
LYSIS A (TYROSINE-PROTEIN KINASE 1) >gi|974334 (U32174)
non-receptor tyrosine kinase (Dictyostelium discoideum)
Length = 1584

Score = 46.5 bits (108), Expect = 2e-04
Identities = 29/106 (27%), Positives = 48/106 (44%), Gaps = 4/106 (3%)

Query: 130 HNEDTTSNTDTSNQATSLDNSTGMTANRNAYVSLPQSEVNID---VDNTTLRFADN-N 185
+NED +SN+ +N N+ +N+ N N+ + N+ ++NTT N N
Sbjct: 442 NNEDISSNNSSNTNNNNNNNNNNNNNSN 501

Query: 186 TIDNGKTVNKSSNESQNAKRQKQGNAGTQFTKQYLIDNIDKA 231
+N N +SN +N N N N N TK+ I + D++
Sbjct: 502 NNNNNNSNSNSNSNNNNNNNNNNNNNNNNNNNNNNIYLTKKPSIGSTDES 547

Score = 34.0 bits (76), Expect = 1.1
Identities = 20/117 (17%), Positives = 46/117 (39%)

Query: 87 NRQTVAPGMQVITVCITHEDYLNVVYSSEVEKYLSQGFTEHNEDTTSNTDTSNQNA 146
N G IT T + + + + + + +N+ +N+ +N N

300

Sbjct: 415 MNNNNNIIGNGKITTITTTTSTSPSSINNEDISSNNNNNNNNNNNNNNNNNNNN 474

Query: 147 TSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTIDNGKTVNKSSNESQN 203
+ + + + T N N + + N + + N + N + N + N + N + N + N

Sbjct: 475 NNNNSNSSNTNNNNINNTTNNNSNSNNNNNNNSNSNSNNNNNNNNNNNNNNNN 531

Score = 33.2 bits (74), Expect = 1.8
Identities = 18/88 (20%), Positives = 35/88 (39%)

Query: 130 HNEDTTSNTDETSQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTIDN 189
+N + +N + +N N T T + S+ +E +N +NN +N

Sbjct: 405 NNNNSNNNNNNNNNNIIGNGKITTITTTTSTSPSSINNEDISSNNNNNNNNNNNNNN 464

Query: 190 GKTIVNKSSNESQNAKRNQKQKNAKGT 217
N +N +N N+ + N T

Sbjct: 465 NNNNNNNNNNNNNNSNSTNNNNINNT 492

Score = 32.5 bits (72), Expect = 3.1
Identities = 18/94 (19%), Positives = 37/94 (39%)

Query: 120 KYLQSQGFTEHNEDTTSNTDETSQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTL 179
K + S N + +N+ +N N ++ + +T S N D+ +

Sbjct: 392 KVVNSTSILVPGNNNNNNNNNNNNNNNNNNNNIIGNGKITTITTTTSTSPSSINNEDISSNN 451

Query: 180 RFADNNTIDNGKTVNKSSNESQNAKRNQKQKGN 213
+NN +N N +N +N N + + N

Sbjct: 452 NNNNNNNNNNNNNNNNNNNNNNNNSNSTN 485

Score = 32.5 bits (72), Expect = 3.1
Identities = 24/110 (21%), Positives = 44/110 (39%), Gaps = 10/110 (9%)

Query: 138 TDETSQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTIDNGK----- 191
T T+ + +S+ +N+ +N N + + N + +N +NN N

Sbjct: 429 TTTTSTSPSSINNEDISSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSSNTNNNN 488

Query: 192 ---TVNKSSNESQNAKRNQKQKNAKGTQFTKQYLIDNIDKAYDLRKK 237
T N +SN +N N N N N+ +N + L KK

Sbjct: 489 INNTTNNNSNSNNNNNNNSNSNSNNNNNNNNNNNNNNNNNNNNIYLTCK 538

>gi|3758855|emb|CAB11140.1| (Z98551) predicted using hexExon;
MAL3P6.11 (PFC0760c), Hypothetical protein, len: 3395 aa
[Plasmodium falciparum]
Length = 3394

Score = 46.5 bits (108), Expect = 2e-04
Identities = 52/202 (25%), Positives = 96/202 (46%), Gaps = 32/202 (15%)

Query: 21 FNEFVNDNKLTIFYDDEFQFMQKMLKFD-KDVLAIVNEKVFGPSLKDELSDL--LFKSF 77
F ++ ++ K T D+ M+K K D DV + NEK++ L ++L+ + + KK

Sbjct: 665 FEKYCSNINKTLIRDD---MKKFRKPDISDVHILHNEKIYLEKLLNEKLNVIKDIKKLD 721

Query: 78 TIHFLDREINRQTVEAFGMQV-----ITVCITHEDVLFVYSSSEVEKYLQSQGFTEHNE 132
+H + IN+ + + +QV I V + DY + S + + K + +N

Sbjct: 722 ELHGV---INKNKEDIYILQVEKQTLIKVISSVDYTKME-SENHIFKMTTWKMLNNV 777

Query: 133 DTSTNTDETSQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTIDNGKT 192
+SN D +NQN ++N+ + N+N N +++N + N +N

Sbjct: 778 HMSSNKDY-MNQNNQNIENNQNIENNQN-----NQNIEN-----NQNIENNQN 820

Query: 193 VNKSSNESQNAKRNQKQKNA 214
N +N++NQN + NQN + NA

Sbjct: 821 QNNQNNQNNQNNQNNQNNQNNNA 842

Score = 33.6 bits (75), Expect = 1.4
Identities = 46/221 (20%), Positives = 89/221 (39%), Gaps = 37/221 (16%)

Query: 10 DFIKSELIKKGFNEFVNDNKLTIFYDDEFQFMQKMLKFDKDVLAIVNEKVFGPSLKDELS 69
D +K E K N + +L Y + +M+K K + V K SL

Sbjct: 367 DSLKIEYNKSKTNIQQLNEQLVNYKNFIKEMKKYK-----QLVVKNNSLFSITH 416

Query: 70 DLLFKKSFTHFLDREINRQTVEAFGMQVITVCITH---EDVLFVYSSSEVEKYLQSQG 126

301

Sbjct: 417 DFINKNSNIIIRRTSDMKQI---FKYNLDIEHFNQDHLSVIY---IYEILYNTN 468

Query: 127 FTEHNEDTTSNTDETSNQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNT 186
+N D +N D +N N + +N+ N N N + +N +

Sbjct: 469 -DNNNDNDNNNDNNNNNNNNNDNNNNNNNDNNNN-----NNNYNNIMM-----M 512

Query: 187 IDNGKTVNKSSNESNQNAKRNQNQKGNAGTQFTKQYLIDN 227

Sbjct: 513 IENMNSGNHPNSNNLHNYRHNTDENNLSSLKTSFRYKINN 553

Score = 32.8 bits (73), Expect = 2.4
Identities = 28/122 (22%), Positives = 53/122 (42%), Gaps = 2/122 (1%)

Query: 119 EKYLQSQGFTEHNEDTTSNTDETSNQNATSLDNSTGMTANRNAYVSLPQSEVNID-VDNT 177

Sbjct: 2838 E Y S + + + + N + + N + + D N + N N + + + N D + + N
ENYPVSTHYDNNDDINKONINNDNNNNONINDDNNNDNINNDNNNDNINNDNINNDNINND 2897

Query: 178 TLRFADNNTIDNGKTVNKSSNESNQNAKRNQNQKGNAGTGQFTKQYLIDNIDKAYDLRKK 237

Sbjct: 2898 NNNDNNNDNSNNGFVCELSSNINDFNNILNVN-KONFQGINKSNFSTNLSEYNYDAYVK 2956

Query: 238 IL 239

Sbjet: 2957 I+ IV 2958

Score = 32.5 bits (72), Expect = 3.1
Identities = 46/249 (18%), Positives = 101/249 (40%), Gaps = 31/249 (12%)

Query: 9 YDFIKSELIKKGFNEFVNDNKLTFYDDEFQFMQKMLKPKDKOVLAIIVNEKVFKGFSLKDEL 68

Sbjct: 2150 YNYVK---VQNATNRDNKNK-----ERNLSQEIYKYINENIDLTSELEKQNDMLENYK 2200

Query: 69 SDL-----LFKKSFTIHFLDREINRQTVEAFGMQVITVCITHEDYLNVVYSSSEVEKYL 122

++L ++K + I L + M+ + + N + B+ + L
Sbjct: 2201 NELKENEBIYKLANDIDMLSNCKKLKESIMMMEKYKIIMN-----NNIQKDEIIENL 2255

Query: 123 QSQGFTEHNEDTTSNTDETSNQNATSLDNSTGMTAN----RNAYVSLPQSE----VNIDV 174

+++ + +D +N + ++S M+ + N + +L +S N+D+
 Sbjct: 2256 KNK-YNNKLDLINSVVDKSI VSCFEDSNIMSPSCNDILNVFNLSKSNKVKCTNM DI 2314

Query: 175 DNTTLRFADNNTIDNGKTVNKSSNESNQNAKRNQNQKGNAGTQFTKQYLIDNIDKAYDL 234

N + ++I+N +N +N +N N N N N K YL++N+ D
 Sbjct: 2315 CNEHMSI--SSINNVNNINNVNNINNVNNINNVNNINNVKNIVDINNYLVNQLQNKDN 2372

Query: 235 RKKILNEFD 243

I+ +P+
Sbjct: 2373 DNIIIIKFN 2381

Score = 32.1 bits (71), Expect = 4.1
Identities = 20/103 (19%), Positives = 48/103 (46%), Gaps = 2/103 (1%)

Query: 115 SSEVEKYLQSQGFTEHNEDTTSNTDETSNQN--ATSLDNSTGMTANRNAYVSLPQSEVNI 172

+++ EKY EH + N D +N+N L ++ ++ + N S ++E+
Sbjct: 3264 NNDEBKYSCHDDKNEHTNNDLLNIDHDNNKNNITDELYSTYNVSVSHNKPDSNKENEIQN 3323

Query: 173 DVDNTTLRFADNNTIDNGKTVNKSSNESNQNAKRNQKGNAX 215

Sbjct: 3324 LISIDSSNENDENDENDENDENDENDENDENDENDENDENDENDENDENDENDK 3366

Score = 30.9 bits (68), Expect = 9.2
Identities = 27/118 (22%), Positives = 53/118 (44%), Gaps = 15/118 (12%)

Query: 104 THEDYLNVVYSSSEV---EKYLQSQGFTEHNEDTTSNTDETSNQNATSLDNSTGMTANR 159

T+ D LN+ + +++ E Y HN+D ++ +E QN S+D+S N
 Sbict: 3280 TNNDLLNIDHDNNKNITDELSTYNSVSHNKOPSNKENEI--QNLISIDSSNENDEND 3337

Query: 160 NAYVSLPOSEVNIDVDNTTLRFADNNTIDNGKTVNKSSNESNQNAKRQKQKGNAGT 217

+++ N + D D N ++ N +E+++N + ++N N +GT
 Sbjct: 3338 EN----DENDENDENDEN-----DENDENDENDENDEKDENDENDENDENFDNNNEG 3386

302

>gi|585795|sp|P21538|REB1_YEAST DNA-BINDING PROTEIN REB1 (QBP)
 >gi|626139|pir||S45907 DNA-binding protein REB1 - yeast
 (Saccharomyces cerevisiae) >gi|536280|emb|CAA84992|
 (Z35918) ORF YBR049c (Saccharomyces cerevisiae)
 >gi|559944|emb|CAA86391| (Z46260) REB1 DNA-binding
 protein (Saccharomyces cerevisiae)
 Length = 810

Score = 45.7 bits (106), Expect = 3e-04
 Identities = 34/158 (21%), Positives = 72/158 (45%), Gaps = 14/158 (8%)

Query: 83 DREINRQTVEAFGMQVITVCITHEDYLNVVYSSSEVEKYLQSQGFTEHNEDTTSNTDETS 142
 D+ N+++VE ++ + V + H+++ +++ K+ + Q E + D N ++ S
 Sbjct: 7 DKNANQESVEEAVLK YVGVGLDHQNHDPQLHTKDLENKHSKKQNI VESSSDVDVNNDDSS 66

Query: 143 NQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTID---NGKTVNKSSNE 199
 N+N + D+S ++A L +E + +VD+ N +D N+ +E
 Sbjct: 67 NRNEDNDDSENISA-----LNANESSSNVDHANSNEQHNAVMDWYLRQTAHNQQDDE 119

Query: 200 SNQNAKRNNQKGNAGTQFTKQYLIDNIDKAYDLRKK 237
 ++N N GN F++ ++ +D D KK
 Sbjct: 120 DDEN--NNNTDNGNDSNNHFSQSDIV--VDDDDDKKK 153

>gi|172372 (M58728) DNA-binding protein (Saccharomyces cerevisiae)
 Length = 809

Score = 45.7 bits (106), Expect = 3e-04
 Identities = 34/158 (21%), Positives = 72/158 (45%), Gaps = 14/158 (8%)

Query: 83 DREINRQTVEAFGMQVITVCITHEDYLNVVYSSSEVEKYLQSQGFTEHNEDTTSNTDETS 142
 D+ N+++VE ++ + V + H+++ +++ K+ + Q E + D N ++ S
 Sbjct: 7 DKNANQESVEEAVLK YVGVGLDHQNHDPQLHTKDLENKHSKKQNI VESSNDVDVNNDDSS 66

Query: 143 NQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTID---NGKTVNKSSNE 199
 N+N + D+S ++A L +E + +VD+ N +D N+ +E
 Sbjct: 67 NRNEDNDDSENISA-----LNANESSSNVDHANSNEQHNAVMDWYLRQTAHNQQDDE 119

Query: 200 SNQNAKRNNQKGNAGTQFTKQYLIDNIDKAYDLRKK 237
 ++N N GN F++ ++ +D D KK
 Sbjct: 120 DDEN--NNNTDNGNDSNNHFSQSDIV--VDDDDDKKK 153

>gi|2952545 (AF051898) coronin binding protein (Dictyostelium
 discoideum)
 Length = 560

Score = 44.9 bits (104), Expect = 6e-04
 Identities = 26/83 (31%), Positives = 39/83 (46%), Gaps = 5/83 (6%)

Query: 131 NEDTTSNTDETSNQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTIDNG 190
 N + +N +N N+ S +NS +N N+ + P N D DN T +NNT +N
 Sbjct: 404 NNNNNNNIINNNSNSNSNNNSNN-NSNNNSNRNSPNHNNNGDNDNNT----NNNTNNNN 458

Query: 191 KTVNKSSNESNQNKRNNQKGN 213
 N ++N +N N N N N
 Sbjct: 459 NNNNNNNNNNNNNNNNNNNNNNN 481

Score = 41.4 bits (95), Expect = 0.006
 Identities = 22/88 (25%), Positives = 43/88 (48%), Gaps = 6/88 (6%)

Query: 130 HNEDTTSNTDETSNQNATSLDN---STGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNT 186
 + ++ +N++ SN N+ + +N + G AN++ + P + +N + DN +NN
 Sbjct: 337 NRNNSNNNSNNNSNNNSNNNSNNRNITNGSNANKS---NSPNNNLATNNDKNKNNNSNNNNN 393

Query: 187 IDNGKTVNKSSNESNQNKRNNQKGN 214
 +N S+N +N N N N N+
 Sbjct: 394 SNNNSNNGNSNNNNNNNNIINNNSNSNS 421

Score = 40.6 bits (93), Expect = 0.011
 Identities = 24/101 (23%), Positives = 41/101 (39%), Gaps = 2/101 (1%)

Query: 115 SSEVEKYLQSQGFTEHNEDTTSNTDETSNQNATSLDNSTGMTANRNAYVSLPQSEVNIDV 174
 S+ L + ++N +N ++ N S +N+ N N S + N +

Sbjct: 370 SNSPNNNLNTNNDNKNNNSNNNNSNNGNSNNTNNNNNIINNNSNSNSNNSNNS 429

Query: 175 DNTTLRFADN--NTIDNGKTVNKSSNESNQAKRNQKQGN 213
+N + R + N N DN N ++N +N N N N N

Subjct: 430 NNNNSNRNSPNHNNNGDNDNNTNNNTNNNNNNNNNNNNNNNNNN 470

Score = 40.2 bits (92), Expect = 0.014
Identities = 21/80 (26%), Positives = 39/80 (48%), Gaps = 9/80 (11%)

Query: 130 HNEDTTSNTDETSNQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTIDN 189
 +N D +NT+ +N N + +N+ N N N + +N +ADN+ **

Sbjct: 442 NNGDNDNNTNNNTNNNNNNNNNNNNNNNNNNNNNN--NNNNNNNNNNNYADNSNNNS 492

Query: 190 GKT VNKSSNESNQNAKRNQN 209
+ N +SN +N N +N+N

Sbjct: 493 SNSNNNSNSNNNDNKNEN 512

Score = 39.5 bits (90), Expect = 0.024
Identities = 26/111 (23%), Positives = 44/111 (39%), Gaps = 20/111 (18%)

Query: 112 VYSSSEVEKYLSQ--GFTEHNEDTTSNTDETSQNATSLDNSTGMTANRNAYVSLPQSE 169

Subjct: 296 VYCTHHHTKFYETHRGLLNNNNSNNNSNSNSNNNNNGINNRNNSNNNSN----- 346

Query: 170 VNIDVDNTTLRFADNNTIDNGKTVNKSS-----NESNQNAKRNQKNGNA 214
+ N ++N I NG NKS+ N +N N N N N +

Sbjct: 347 ---NNSNNNSNNSNNRNITNGSNANKSNSPNNNLNTNNDNKNNNSNNNNS 394

Score = 37.5 bits (85), Expect = 0.094
Identities = 24/96 (25%), Positives = 41/96 (42%), Gaps = 1/96 (1%)

Query: 124 SQGFTEHNE~~DTT~~SN~~TD~~ETSNQ~~NAT~~SLDNSTGM-TANRNA~~YV~~SLPQSEVNIDVDNTTLRPA 182

[illegible]

Query: 183 DNNTIDNGKTVNKSSNESNQNAKRNQKGNAGTQ 218
+NN DN + +SN +N N+ N + K O

Sbjct: 481 NNNYADNSNNNS SNSNNNSNSNNNDNKNNSDNQ 516

Score = 35.6 bits (80), Expect = 0.36
Identities = 25/99 (25%), Positives = 42/99 (42%), Gaps = 18/99 (18%)

Query: 130 HNEDTTSNTDETSNQNATSLDNST-GMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTID 188
 +N + SN + +N N ++ N T G AN++ + P + +N + DN +NN +

Subject: 339 NNSNNNSNNNSNNNSNNNSNNNRNITNGSNANKS---NSPNNNLNTNNDNKNNNSNNNNNSN 395

Query: 189 NGKTV-----NKSSNESNQNAKRNQNQKGN 213
N S++ SN N+ N N N

Subject: 396 NNSNNGNSNNNNNNNNIINNNSNSNSNNSNNSNNSN 434

Score = 35.2 bits (79), Expect = 0.47
Identities = 21/94 (22%), Positives = 42/94 (44%), Gaps = 5/94 (5%)

Query: 124 SQGFTEHNEEDTTSNTDETSNQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFAD 183
+ G + ++ +N T+N N + N+ N N+ + N + +N + +

Subject: 362 TNGSNANKSNSPNNLNTNNDNKNNNSNN-----NNNSNNNSNNGNSNNNNNNNNIINN 416

Query: 184 NNTIDNGKTVNKSSNESNQNAKRNQNQKGNAGT 217
+N+ N + N S+N SN+N+ + N N T

Sbjct: 417 SNSNSNNNSNNNSNNNSNRNSPNHNNNGDNDNNT 450

Score = 35.2 bits (79), Expect = 0.47
Identities = 29/118 (24%), Positives = 53/118 (44%), Gaps = 12/118 (10%)

Query: 115 SSEVEKYLS-QGFTEHNEDTTSNTDETSNQNATSLDNSTGMTANRNAYVSLPQSEVNID 173
SS+ R ++ +GF+ + T+N ++N D S+G + + + V+ P+S +N

Sbict: 114 SSDSEADIEDDKGFOD--KPITTNNSGSNNPLKNLKDYSSGSSGSSRSGVNQPRSNINNS 171

Query: 174 VDNTTLRFADNNT-----IDNGKTVNKSSNESNQNAKRNNQNGNAKGTOFTKQ 222

304

Subject: 172 NDKYKSKSSSSNSNSSSSSGGSLISSLLTGGNTYQNQNQNQNQNQNQNNNQSQLOQQQQ 229

Score = 34.4 bits (77), Expect = 0.81
Identities = 24/94 (25%), Positives = 38/94 (39%), Gaps = 12/94 (12%)

Query: 131 NEDTTSNTDETSNQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTITDNG 190
N + T + N + + N N + + N+ N N S N N + NN+ N

Sbjct: 451 NNNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYADNSNNSSNSN-----NNNSNSNN 504

Query: 191 KTVNKSSNESNQNAKR-----NQNQKGNAGTQT 218
 NK+ N NQ+ R ++NQK + Q
 Sbjct: 505 NNDNKNKENSNDNOSVLRSNKFTDENQKNGSDDQQ 538

Score = 33.6 bits (75), Expect = 1.4
Identities = 22/90 (24%), Positives = 35/90 (38%)

Query: 124 SQGFTEHNEDTTSNTDTSNQNATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFAD 183
S N SN +++++ N+ N N + + N + +N
Shift: 353 SNNNNNNNITGSGNANKNSNNNNLNTNNNNKNNNSNNNNNSNNNSNNNGNSNNNNNNNNII 412

Query: 184 NNTIDNGKTVNKSSNESNQNAKRQKQKGN 213
 NN N + N S+N SN N+ RN N
 Sbjct: 413 NNNNSNSNSNNNSNNNSNNNSNRNSPNNNN 442

```
>gi|535260|emb|CAA82996| (Z30339) STARP antigen (Plasmodium  
reichenowi)  
Length = 655
```

Score = 44.5 bits (103), Expect = 7e-04
Identities = 31/114 (27%), Positives = 47/114 (41%), Gaps = 14/114 (12%)

Query: 128 TEHNEDTTSNTDETSQNQATSLDNSTGTMANRNAYVSLPQSEVN-----IDVDNTTLRF 181
T++N T TD + + +N+T A N + ++ N D +NT +
Subject: 433 TDNNNTNTKATDSNNNTNTKATDNNNTNTKATDNNNTNTKATDNNNTNTKATDNNNTNTKA 492

Query: 182 ADNNTI-----DNGKTVNKSSNESNQNAKRNQKGNAGT---QFTKQYLIDN 227
 DNN DN T K+++ +N N K N N K T T QY+ N
 Subject: 493 TDNNNTNTKATDNNNTNTKATDNNNTNTKATDNNNTNTKATDNNNTNTQYVFN 546

Score = 44.5 bits (103), Expect = 7e-04
Identities = 30/103 (29%), Positives = 44/103 (42%), Gaps = 13/103 (12%)

Query: 128 TEHNEDTTSNTDETSNQNATSLDNS----TGMTANRNAYVSLPQSEVN----IDVNTTL 179
T++N T TD++N + + DN+ T T N N S D +NT
Sbjct: 401 TDNNNTDTKATDKSNNTDTKATDNNNTDTKATDNNNTNTKATDSNNNTNTKATDNNNTNT 460

Query: 180 RFADNNNTI-----DNGKTVNKSSNESQNAKRNQKGNAGKT 217
+ DNN DN T K+++ +N N K N N K T
Sbjct: 461 KATDNNNTNTKATDNNNTNTKATDNNNTNTKATDNNNTNTKAT 503

Score = 42.6 bits (98), Expect = 0.003
Identities = 27/96 (28%), Positives = 43/96 (44%), Gaps = 10/96 (10%)

Query: 128 TEHNEDTTSNTDETSNQNATSLD-NSTGMTANRNAYVSLPQSEVNIIDVNTTLRFADNNT 186
T++N +T + + +N N + D N+T A N + ++ N NT + DNN
Subject: 422 TDNNNNNTDKATDNNNTNTKATDSNNTNTKATDNNNTNTKATDNN---NTNTKATDNNN 477

Query: 187 I-----DNGKTVNKSSNESNQNAKRNQKGNQKNAKGT 217
DN T K+++ +N N K N N K T
Subject: 478 TATKATDNNNTATKATDNNNTATKATDNNNTATKAT 513

Score = 41.8 bits (96), Expect = 0.005
Identities = 35/150 (23%), Positives = 59/150 (39%), Gaps = 9/150 (6%)

Query: 85 EINRQTVEAFGMQVITVCITHEDYLVNVYSSSEVEKYLSQGFTEHNEDTTSNTDETSQ 144
E N + + + G T + + N + E + + Q T + N T T + + N
Subject: 118 ETNKTNIKLTGNSTTINTNLIENTNA--TKKLTENVITNOILTGNNTTTNTTSSTEHN 175

Query: 145 NATSLDNSTGMTANRNAYVSLPQSEVNIDVDNTTLRFADNNTIDNGKTVNKSSNESNQNA 204
N + NSTG T+ NI + N L +N T + T + ++ +N N+

305

Sbjct: 176 NINTNTNSTGNTSTTKKLTE-----NI-ITNQILTGNNNTTTNTSSTEHNNTNTNTNS 228

Query: 205 KRNNQKGNAGTQPTKQYLIDNIDKAYDL 234

N N N T + DNI+ +L

Sbjct: 229 TDNSNTNTNLTDITTTTKKWDNINTTQNL 258

Score = 41.8 bits (96), Expect = 0.005
Identities = 30/101 (29%), Positives = 43/101 (41%), Gaps = 13/101 (12%)

Query: 130 HNEDETTSTNDETSNQATSLDNTGTMANRNAYVSLPQSEVNIDV-----DNTTLRFA 182
+N DT S ++ ++ AT DN+ T T N N + N D +NT +

Sbjct: 363 NNITDITSTNDNTDKATDNDNTDKATDNNNTDITDKATDNNNTDITDKATDKSNNITDKAT 422

Query: 183 DNN-----TIDNGKTVNKSSNESNQAKRNQKGNAGT 217
DNN DN T K+++ +N N K N N K T

Sbjct: 423 DNNNTDITDKATDNNNTDKATDNNNTDKATDNNNTDKATDNNNTDKAT 463

Score = 40.6 bits (93), Expect = 0.011
Identities = 31/121 (25%), Positives = 47/121 (38%), Gaps = 31/121 (25%)

Query: 128 TEHNEDTTSTNDETSNQAT-----SLDNTGTMANRNAYVSLPQSEVN----- 171
TEHN + +NT+ T N + T ++ + +T N N + +E N

Sbjct: 171 TEHNNTNTNSTGNTSTTKKLTENIITNQILTGNNNTTTNTSSTEHNNTNTNTNSTD 230

Query: 172 -----IDVDNTTLRFADN-----NTIDNGKTVNKSSNESNQAKRNQKGNAG 216
D+ TT ++ DN T N TV+ +N +N N K N N K

Sbjct: 231 NSNTNTNLTDITTTTKKWDNINTTQNLTTSTNTTSTVSTONNNNTKPTDNNNTNKS 290

Query: 217 T 217

T

Sbjct: 291 T 291

Score = 38.3 bits (87), Expect = 0.055
Identities = 28/98 (28%), Positives = 41/98 (41%), Gaps = 10/98 (10%)

Query: 128 TEHNEDTTSTNDETSNQATSLDNTGTMANRNAYVSLPQSEVNIDVD-NTTLRFADNNT 186
TEHN + +NT+ S N+ + N T +T + N+ NTT DNN

Sbjct: 216 TEHNNTNTNTN--STDNSNTNTNLTDITTTTKKWDNINTTQNLTTSTNTTSTVSTONNN 273

Query: 187 -----IDNGKTVNKSSNESNQAKRNQKGNAGT 217

DN T KS++ N K N+ + K T

Sbjct: 274 NNINTKPTDNNNTNKSNTDNTGTGKETDNKNTDIKAT 311

Score = 37.5 bits (85), Expect = 0.094
Identities = 31/106 (29%), Positives = 45/106 (42%), Gaps = 18/106 (16%)

Query: 128 TEHNEDTTSTNDETSNQATSLDNTGTMANRNAYVSLPQSEVN-----IDVDN 176
T++N +T +T T N N AT N+T A N + ++ N D +N

Sbjct: 390 TDNNNT--DTKATDNNNTDKATDKSNNITDKATDNNNTDITDKATDNNNTDKATDSNN 447

Query: 177 TILRFADNN-----TIDNGKTVNKSSNESNQAKRNQKGNAGT 217

T + DNN DN T K+++ +N N K N N K T

Sbjct: 448 TMTKATDNNNTDKATDNNNTDKATDNNNTDKATDNNNTDKATDNNNTDKAT 493

Score = 35.2 bits (79), Expect = 0.47
Identities = 24/109 (22%), Positives = 46/109 (42%), Gaps = 6/109 (5%)

Query: 128 TEHNEDTTSTNDETSNQATSLDNTGTMANRNAYVSLPQSEVN-----IDVDNTTLRF 181
T++N T TD + + +N+T A N + ++ N D +NT +

Sbjct: 473 TDNNNTNTKATDNNNTNTKATDNNNTNTKATDNNNTNTKATDNNNTNTKATDNNNTNTKA 532

Query: 182 ADNNTIDNGKTVNKSSNESNQAKRNQKGNAGTQPTKQYLIDNIDK 230

DNN N + +E+ + K N++ N++ + K + +DK

Sbjct: 533 TDNNNTNQYVFANNYDETTSDDKLNKDCDNNSEKENIKSMINAYLDK 581

Score = 34.4 bits (77), Expect = 0.81
Identities = 26/126 (20%), Positives = 46/126 (35%), Gaps = 7/126 (5%)

306

Query: 99 ITVCITHEDYLNVVYSSSEVEKYLSQSGFTEHNEDTTSNTDETSNQNATSLDNSTGMTAN 158
 IT T+ + ++ S + V S T +++ +N T N N ++ T
 Sbjct: 318 ITDNTNTNVISTDNSKTNVISKDNSNTHISTDNSKTNVISTDNNNTDTISTDNDNTDT 377

Query: 159 RNAYVSLPQSEVNIDVDNTTLRFADNNTID-----NGKTVNKSSNESNQNAKRQK 211
 + + + +NT + DNN D N + N +N + K N
 Sbjct: 378 KATDNDNTDTKATDNNNTDTKATDNNNTDTKATDKSNNTDTKATDNNNTDTKATDNN 437

Query: 212 GNAKGT 217
 N K T
 Sbjct: 438 TNTKAT 443

Score = 34.4 bits (77), Expect = 0.81
 Identities = 30/100 (30%), Positives = 44/100 (44%), Gaps = 14/100 (14%)

Query: 131 NEDTTSNTDETSNQNATSLDNTS-TGMTANRNAY---VSLPQSEVNI---DVDNTTLRFAD 183
 N + T TD T N N S DNS T + + N+ +S S+ N+ D +NT D
 Sbjct: 313 NNNITITDNT-NTNVISTDNSKTNVISKDNSNTHISTDNSKTNVISTDNNNTDTISTD 371

Query: 184 NNTIDNGKTVNKSS-----NESNQNAKRQKQKGNAGT 217
 N+ D T N ++ N +N + K N + K T
 Sbjct: 372 NDNTDTKATDNDNTDTKATDNNNTDTKATDNNNTDTKAT 411

Score = 34.4 bits (77), Expect = 0.81
 Identities = 28/101 (27%), Positives = 41/101 (39%), Gaps = 15/101 (14%)

Query: 131 NEDTTSNTDETSNQNATSLDNTS-TGMTANRNAYVSLPQSEVNIDV-----DNTTLRFA 182
 N DT + ++ ++ AT +N+T A N N N D +NT +
 Sbjct: 374 NTDTKATDNDNTDTKATDNNNTDTKATDNNNTDTKATDKSNNTDTKATDNNNTDTKAT 433

Query: 183 DNNTIDNGK-----TVNKSSNESNQNAKRQKQKGNAGT 217
 DNN N K T K+++ +N N K N N K T
 Sbjct: 434 DNNN-TNTKATDSNNTNTKATDNNNTNTKATDNNNTNTKAT 473

Score = 32.5 bits (72), Expect = 3.1
 Identities = 30/110 (27%), Positives = 40/110 (36%), Gaps = 23/110 (20%)

Query: 131 NEDTTSNTDETSNQNATSLDNTS-----TGMTANRNAYVSLPQS-----EVNIDVDNTTLRF 181
 N +TT N ++N S DN+ T T N N + + D NT ++
 Sbjct: 251 NINTTQNLTTSTNTTTVSTDNNNNNINTKPTDNNNTNIKSTDNYNTGKTEDNKNNTDIKA 310

Query: 182 ADNNTI-----DNGKTVNKSSNESNQNAKRQKQKGNAGT 217
 DNN I DN KT S + SN + N K N T
 Sbjct: 311 TDNNNTITDNTNTNVISTDNSKTNVISKDNSNTHISTDNSKTNVIST 360

>gi|1429240|emb|CAA67659| (X99260) lower collar protein
 [Bacteriophage B103]
 Length = 293

Score = 43.8 bits (101), Expect = 0.001
 Identities = 53/204 (25%), Positives = 79/204 (37%), Gaps = 42/204 (20%)

Query: 56 EKVPKG----FSLKDELSDLLPKKSFTIHFLD----REINRQTVEAFGMQVITVCITHED 107
 EK+ KG P + + D ++K F HF+ REI +T P + T I +
 Sbjct: 26 EKIEKGRPKLFDPPQYPIFDESRYKVPETHFIRNFYMEIGFETEGLEKFNLETWLIINMP 85

Query: 108 YLNVVYSSSEVEKY-----LQSQGFTEH-----NEDT-----SNTDETSNQNA 146
 Y N ++ S E+ KY L + G ++ N DTT SNT + NA
 Sbjct: 86 YFNKLPES-ELIKYDPLENTRLNNTGNKKNDTERNDNRDTTGSMKADGKSNTTKSDKTNA 144

Query: 147 TSLDNTS-TGMTA-----NRNAYVSLPQSEVNIDVDN--TTLRFADNNTIDNGKTVNKS 196
 T G T NR P S +N+ ++ TL +A + I+ T NK
 Sbjct: 145 TGSSKEDGKTTGVSVDNPNKIDSDQPOSRLNLTNDGQGTLEYA--SAIEENNTNNKR 202

Query: 197 SNESNQNAKRQKQKGNAGTQFT 220
 + N + + GT T
 Sbjct: 203 NTTGTNNVTSSAESESTGSGTSDI 226

Query= pt|110879 44AHJDORF009 Phage 44AHJD ORF |5744-6496|2 1
 (250 letters)

307

>gi|2764981|emb|CAA69021.1| (Y07739) N-acetylmuramoyl-L-alanine
amidase [Staphylococcus phage Twort]
Length = 467

Score = 180 bits (452), Expect = 1e-44
Identities = 89/157 (56%), Positives = 109/157 (68%), Gaps = 8/157 (5%)

Query: 1 MKSQQAKKEWIKHEGAGVDFDGYGFCMDLSVAYVYITDGKVRMGNKADAINNDFK 60
MK+ +QA+ +I G DFDG YG+QCMDL+V Y+Y++TDGK+RMWGNKADAINN F
Sbjct: 1 MXTLKQAESYIKSVNTGTDGGLYGYQCMDLAVDYIYHVTDGKIRMWGNKADAINNSFG 60

Query: 61 GLATVYKNTSPFKPQLGDVAVYTNGQ---YGHICVLS----GNLDYITCLEQNWLGGGF 113
G ATVYKN P+P+P+ GDV V+T G YGHI V + G+L Y T LEQNW G G
Sbjct: 61 GTATVYKNYPFRPKYGDVVVMTGNFATYGHIAIVTNPDPYGLQYVTVLEQNWNGNGI 120

Query: 114 DGWEKATIRTHYTDGVTHFIRPKFSGSNS-KALETSK 149
E ATIRTH Y G+THFIRP F+ +S K +T K
Sbjct: 121 YKTELATIRTHDYTGITHFIRPNFATESSVKQKDTKK 157

Score = 61.7 bits (147), Expect = 6e-09
Identities = 41/125 (32%), Positives = 57/125 (44%), Gaps = 8/125 (6%)

Query: 125 YYDGVTHFIRPKFSGSNSKALETSKVNTFGKWKRNQYGYTYRNENGTFTC-GPLPIFARV 183
YY+G T P +K + +T G W N YGTY+ +E+ TF C I R
Sbjct: 346 YYEKTPV--PTVVNQKAKTKPKVQSSTSG-WNVNNYGYTYKSESATPKCTARQIVTRY 402

Query: 184 GSPKLEPNGTWFPNGYTPYNEVCLSDGYVWIGYNWQTR-YYLPVRQWNGKTGNSYSV 242
P + P Y+ VC DGYVWI + G + ++PVR W+ N+ +
Sbjct: 403 TGFPTTCQAGVLYYQSVTYDTVCKQDGYVWISWITNGGQDVWMPVRTWD---KNTDIM 459

Query: 243 GIPWG 247
G WG
Sbjct: 460 GQLWG 464

>gi|113675|sp|P24556|ALYS_STAAU AUTOLYSIN
(N-ACETYLMURAMOYL-L-ALANINE AMIDASE)
>gi|79887|pir|JQ1147 N-acetylmuramoyl-L-alanine amidase
(EC 3.5.1.28) - Staphylococcus aureus >gi|153067
(M76714) peptidoglycan hydrolase [Staphylococcus aureus]
Length = 481

Score = 118 bits (292), Expect = 6e-26
Identities = 56/117 (47%), Positives = 68/117 (57%), Gaps = 1/117 (0%)

Query: 135 PKFSGSNSKALETSKVNTFGK-WKRNQYGYTYRNENGTFTCGFLPIFARVGSPLSEPN 193
P + SN + ++ V WKRN+YGTY E+ FT G PI R P LS P G
Sbjct: 365 PVATVSNSSASSNTVKPVASAWKRNKYGYTYMEESARFTNGNQIPITVRKVGPFLLSCPVG 424

Query: 194 YWFQPNQYTPYNEVCLSDGYVWIGYNWQTRYLLPVRQWNGKTGNSYSVGIPIHGVFS 250
Y FQF GY Y EV L DG+VW+GY W+G RYTLF+R WNG + +G NG S
Sbjct: 425 YQFPQGGYCDYTEVMLQDGHVWVGTYWEGQRYTLPIRTWNGSAPPNQILGDLNGEIS 481

Score = 78.0 bits (189), Expect = 7e-14
Identities = 48/109 (44%), Positives = 62/109 (56%), Gaps = 6/109 (5%)

Query: 15 EGAGVDFDGYGFCMDLSVAYVYITDGKVRMGNKADA-INNDFKGLATVYKNTSPFK 73
EG + D YGFQC D + A + + G + AKD N+F GLATVY+NTP F
Sbjct: 18 EGKQFNVDLWYGFQCFDYANAG-WKVLFGLLKGLGAKDIPFANNFDGLATVYQNTPDFL 76

Query: 74 PQLGDVAVYTNGQ---YGHICVLSGNLDYITCLEQNWLGGGF-DGWEK 118
Q GD+ V+ + YGH+ V+ LDY EQNWLGGS+ DG E+
Sbjct: 77 AQPQDMVVFGSNYAGYGHVAVVIEATLDYIIIVYEQNWLGGGWTGIEQ 125

>gi|1763243 (U72397) amidase [bacteriophage 80 alpha]
Length = 481

Score = 118 bits (292), Expect = 6e-26
Identities = 56/117 (47%), Positives = 68/117 (57%), Gaps = 1/117 (0%)

Query: 135 PKFSGSNSKALETSKVNTFGK-WKRNQYGYTYRNENGTFTCGFLPIFARVGSPLSEPN 193
P + SN + ++ V WKRN+YGTY E+ FT G PI R P LS P G
Sbjct: 365 PVATVSNSSASSNTVKPVASAWKRNKYGYTYMEESARFTNGNQIPITVRKVGPFLLSCPVG 424

Query: 194 YWQPNGYTPYNEVCLSDGYVWIGYNWQGRYYLPVRQWNGKTGNSYSVGIPWGVFS 250
 Y FQP GY Y EV L DG+VW+GY W+G RYYLP+R WNG + +G WG S
 Sbjct: 425 YQFPQGGYCDYTEVMLQDGHVWVGYYTWEGQRYLPIRTWNGSAPPNQILGDLNGEIS 481

Score = 83.5 bits (203), Expect = 2e-15
 Identities = 50/115 (43%), Positives = 65/115 (56%), Gaps = 6/115 (5%)

Query: 9 EWIYKHEGAGVDFDGYGFCMDLSVAYVYYITDGKVRMNGNAKDA-INNDFKGLATVYK 67
 EW+ EG + D YGFQC D + A + + G + AKD N+F GLATVY+
 Sbjct: 12 EWLKTSSEKQFNVDLWYGFQCFDYANAG-WKVLFGLLKGLGAKDIPFANNFDGLATVYQ 70

Query: 68 NTPSFKPQLGDAVYVTNGQ---YGHICVLSGNLDYYTCLEQNWLGSGF-DGWEK 118
 NTP F Q GD+ V+ + YGH+ V+ LDY EQNWLGSG+ DG E+
 Sbjct: 71 NTPDFLAQPGDMVVFSGSYGAGYGHVAVIBATLDYIIVYEQNWLGSGWTDGIEQ 125

>gi|4574237|gb|AAD23962.1|AF106851_1 (AF106851) *LytN* (*Staphylococcus aureus*)
 Length = 383

Score = 84.3 bits (205), Expect = 9e-16
 Identities = 48/128 (37%), Positives = 68/128 (52%), Gaps = 7/128 (5%)

Query: 15 EGAGVDFDGYGFCMDLSVAYVYYITDGKVRMNGNAKDAINNDFKGLATVYKNTPSFKP 74
 E G DFDG+YG+QC DL Y ++ ++ +G N+F A +Y NTP+FK
 Sbjct: 252 ENRGWDFDGSYGNQCFDLVNVYWNHLYGHGLKGYGAKDIPYANNFNSEAKIYHNTPTFKA 311

Query: 75 QLGDVAVYT---NGQYGHICVLSGNLD---YTCLEQNWLGSGFDGWEKATIRTHYYD 127
 + GD+ V++ G YGH VL+G+ D + L+QNW GG+ E A H Y+
 Sbjct: 312 EPGDLVVFSGRFGGGYGHATAVLNGDYDGKLMKFSLDQNWNGGWRKAEVAHKVVHNYE 371

Query: 128 GVTHFIRP 135
 FIRP
 Sbjct: 372 NDMIFIRP 379

>gi|3767593|dbj|BAA33856.1| (AB015195) *LytN* (*Staphylococcus aureus*)
 Length = 383

Score = 84.3 bits (205), Expect = 9e-16
 Identities = 48/128 (37%), Positives = 68/128 (52%), Gaps = 7/128 (5%)

Query: 15 EGAGVDFDGYGFCMDLSVAYVYYITDGKVRMNGNAKDAINNDFKGLATVYKNTPSFKP 74
 E G DFDG+YG+QC DL Y ++ ++ +G N+F A +Y NTP+FK
 Sbjct: 252 ENRGWDFDGSYGNQCFDLVNVYWNHLYGHGLKGYGAKDIPYANNFNSEAKIYHNTPTFKA 311

Query: 75 QLGDVAVYT---NGQYGHICVLSGNLD---YTCLEQNWLGSGFDGWEKATIRTHYYD 127
 + GD+ V++ G YGH VL+G+ D + L+QNW GG+ E A H Y+
 Sbjct: 312 EPGDLVVFSGRFGGGYGHATAVLNGDYDGKLMKFSLDQNWNGGWRKAEVAHKVVHNYE 371

Query: 128 GVTHFIRP 135
 FIRP
 Sbjct: 372 NDMIFIRP 379

>gi|2764983|emb|CAA69022.1| (Y07740) cell wall hydrolase Ply187
 (*Staphylococcus phage 187*)
 Length = 628

Score = 76.9 bits (186), Expect = 2e-13
 Identities = 50/144 (34%), Positives = 68/144 (46%), Gaps = 18/144 (12%)

Query: 5 QQAKEWIYKHEGAGVDFDGYGFCMDLSVAYVYYITDGKVRMW-----GNAKDAINNDF 59
 +Q +W G+GVD DG YG QC DL Y++ R W GNA+D +
 Sbjct: 12 KQVVDWAINLIGSGVDVDGYGRQCWDLP-NYIFN-----RYWNFKTPGNARDMAWYRY 64

Query: 60 KGLATVYKNTPSFKPQLGDAVYVTNGQY-----GHIQCVLS-GNLDYYTCLEQNWLGSGF 113
 V++NT F P+ GD+AV+T G Y GH V+ Y+ ++QNW
 Sbjct: 65 PEGFKVFRNTSDFVFKPGDIAVWTGGNYNWNMTWGHTGIVVGPSTKSYFYSDQNWNNNSNS 124

Query: 114 DGWEKATIRTHYYDCVTHFIRPKF 137
 A H Y GVTHF+RP +
 Sbjct: 125 YVGSAAKIKHSYFGVTHFVRPAY 148

309

>gi|3287732|sp|O05156|ALE1_STACP GLYCYL-GLYCINE ENDOPEPTIDASE ALE-1
 PRECURSOR >gi|1890068|dbj|BAA13069| (D86328) ALE-1
 [Staphylococcus capitis]
 Length = 362

Score = 73.4 bits (177), Expect = 2e-12
 Identities = 47/117 (40%), Positives = 61/117 (51%), Gaps = 10/117 (8%)

Query: 132 FIRPKFSGSNSKALETSKVNTFGKWKRNQYGYTYRNENGTFTCGFLPIFARVGSPLKSEP 191
 F++ GSNS TS N G +K N+YGT Y++E+ +FT I R+ P S P
 Sbjct: 252 FLKSAGYGSNS----TSSNNNG-YKTNKYGTLYKSESASFTAN-TDIITRLTGPFRRSMP 305

Query: 192 NGYWFQPNGYTPYNEVCLSDGYVWIGYNW-QGTRYLPLVRQWNGKTGNSYSVGIPWG 247
 + Y+EV DG+VW+GYN G R YLPVR WN TG +G WG
 Sbjct: 306 QSGVLRKGLTIKYDEVKQDGHVWVGYNVNSGKRVYLPVRTWNESTG---ELGPLWG 359

>gi|79926|pir|A25881 lysostaphin precursor - Staphylococcus
 simulans >gi|153047 (M15686) lysostaphin (ttg start
 codon) [Staphylococcus simulans]
 Length = 389

Score = 69.5 bits (167), Expect = 3e-11
 Identities = 48/133 (36%), Positives = 62/133 (46%), Gaps = 20/133 (15%)

Query: 131 HFIRPKFSGSNSKALETS---KVNTFGK-----WKRNQYGYTYRNENGTFTCG 175
 HF R S SNS A + K +GK WK N+YGT Y++E+ +FT
 Sbjct: 258 HFQRMVNSFSNSTAQDPMPLKSAGYKGAGGTVTPTNTGKTNKYGTLYKSESASFTPN 317

Query: 176 FLPIFARVGSPLKSEPNQYWFQPNQYTPYNEVCLSDGYVWIGYNW-QGTRYLPLVRQWNG 234
 I R P S P + Y+EV DG+VW+GY G R YLPVR WN
 Sbjct: 318 -TDIITRTTGPFRRSMPQSGVLKAGQTIHYDEVKQDGHVWVGTYGNSGQRIYLPVRTWNK 376

Query: 235 KTGNSYSVGIPWG 247
 T ++G+ WG
 Sbjct: 377 STN---TLGVLWG 386

>gi|126496|sp|P10548|LSTP_STAST LYSOSTAPHIN PRECURSOR
 (GLYCYL-GLYCINE ENDOPEPTIDASE) >gi|79927|pir||S01079
 lysostaphin precursor - Staphylococcus simulans bv.
 staphylolyticus >gi|581744|emb|CAA29494| (X06121)
 lysostaphin (AA 1-480) [Staphylococcus simulans bv.
 staphylolyticus]
 Length = 480

Score = 69.5 bits (167), Expect = 3e-11
 Identities = 48/133 (36%), Positives = 62/133 (46%), Gaps = 20/133 (15%)

Query: 131 HFIRPKFSGSNSKALETS---KVNTFGK-----WKRNQYGYTYRNENGTFTCG 175
 HF R S SNS A + K +GK WK N+YGT Y++E+ +FT
 Sbjct: 349 HFQRMVNSFSNSTAQDPMPLKSAGYKGAGGTVTPTNTGKTNKYGTLYKSESASFTPN 408

Query: 176 FLPIFARVGSPLKSEPNQYWFQPNQYTPYNEVCLSDGYVWIGYNW-QGTRYLPLVRQWNG 234
 I R P S P + Y+EV DG+VW+GY G R YLPVR WN
 Sbjct: 409 -TDIITRTTGPFRRSMPQSGVLKAGQTIHYDEVKQDGHVWVGTYGNSGQRIYLPVRTWNK 467

Query: 235 KTGNSYSVGIPWG 247
 T ++G+ WG
 Sbjct: 468 STN---TLGVLWG 477

>gi|3287967|sp|P10547|LSTP_STASI LYSOSTAPHIN PRECURSOR
 (GLYCYL-GLYCINE ENDOPEPTIDASE) >gi|2072411 (U66883)
 lysostaphin [Staphylococcus simulans]
 Length = 493

Score = 69.5 bits (167), Expect = 3e-11
 Identities = 48/133 (36%), Positives = 62/133 (46%), Gaps = 20/133 (15%)

Query: 131 HFIRPKFSGSNSKALETS---KVNTFGK-----WKRNQYGYTYRNENGTFTCG 175
 HF R S SNS A + K +GK WK N+YGT Y++E+ +FT
 Sbjct: 362 HFQRMVNSFSNSTAQDPMPLKSAGYKGAGGTVTPTNTGKTNKYGTLYKSESASFTPN 421

Query: 176 FLPIFARVGSPLKSEPNQYWFQPNQYTPYNEVCLSDGYVWIGYNW-QGTRYLPLVRQWNG 234

310

I R P S P + Y+EV DG+VW+GY G R YLPVR WN
 Sbjct: 422 -TDIITRTTGPFRSMPQSGVLKAGQTIHYDEVMMKQDGHVWVGTYGNSGQRIYLPVRTWNK 480

Query: 235 KTGNSYSVSGIPWG 247

T ++G+ WG

Sbjct: 481 STN---TLGVWLG 490

>gi|3341932|dbj|BAA31898.1| (AB009866) amidase (peptidoglycan
 hydrolase) [bacteriophage phi PVL]
 Length = 484

Score = 68.3 bits (164), Expect = 6e-11
 Identities = 52/150 (34%), Positives = 71/150 (46%), Gaps = 17/150 (11%)

Query: 3 SQQQAKEWIKHEGAGVDFDAGYGFQCMDSVAYVYITDGKVRMWGNADAINNDFKGL 62

++ QA++W G + D YGFQC D + + + I G+ R+ G I D K

Sbjct: 4 TKKQAQKWFQNSLQKQFNPDLFYGFQCYDYASMF-FMIATGE-RLQGLYAYNIPFDNKR 61

Query: 63 ATVY---KNTSPFKPQLGDVAVYTN---GQYGHICVLSGNLDYITCLEQNWLGSGF-- 113

Y KN SP PQ D+ V+ + G GH++ V S NL+ +T QNW G G+

Sbjct: 62 IEYGGQIIKQYDSFLPQKLDIVVFPKSKYGGGAGHVEIVESANLNTPTSFGQNWNGKGTN 121

Query: 114 ----DGW--EKATIRTHYYDGVTHFIRPKF 137

GW E T HYYD +FIR F

Sbjct: 122 GVAQPGWGPETVTRHVHYDDPMYFIRLNF 151

Query= pt|110882 44AHJDORF012 Phage 44AHJD ORF |8391-8813|3 1
 (140 letters)

>gi|140528|sp|P24811|YQXH_BACSU HYPOTHETICAL 15.7 KD PROTEIN IN
 SPOIIC-CWLA INTERGENIC REGION (ORF2)
 >gi|322189|pir||B44816 orf2 5' of autolytic amidase -
 Bacillus subtilis >gi|142801 (M59232) open reading frame
 2 [Bacillus subtilis] >gi|1217874|dbj|BAA06959| (D32216)
 ORF121 [Bacillus subtilis] >gi|1303767|dbj|BAA12423|
 (D84432) Yqdd [Bacillus subtilis]
 >gi|2635036|emb|CAB14532| (Z99117) alternate gene name:
 yqdd; similar to holin [Bacillus subtilis]
 Length = 140

Score = 80.4 bits (195), Expect = 6e-15
 Identities = 45/130 (34%), Positives = 67/130 (50%), Gaps = 3/130 (2%)

Query: 4 VKFRFTDSEAFHMFYAGDLKLLYFLFVLMFVDIITGSKAIKNNLWSKSKMRGFSKXX 63

+ F D ++P G +K L L VL +D++TG+ KA K L S+ + G+ +K

Sbjct: 8 INFETLDLARVYLF---GGVKYLDLLVLSTIIDVLTGVKAWKFKLRSRSANFGYVRKL 64

Query: 64 XXXXXXXXXXXXXXXXXXXXKGLLMITIFYIANEGLSIVENCAEMDVLVPEQIKDKLRVI 123

G L T+ +YIANEGLSI EN A++ V +P I D+L+ I

Sbjct: 65 LNFFAVILANVIDTVLNLNGVLTFTGTVLFYIANEGLSITENLAQIGVKIPSSITDRLQTI 124

Query: 124 KNDTEKSDNN 133

+N+ E+S NN

Sbjct: 125 ENKEQSKNN 134

>gi|4126631|dbj|BAA36651.1| (AB016282) ORF45 [bacteriophage phi-105]
 Length = 135

Score = 76.1 bits (184), Expect = 1e-13
 Identities = 44/115 (38%), Positives = 61/115 (52%), Gaps = 4/115 (3%)

Query: 21 GDLKLLYFLFVLMFVDIITGSKAIKNNLWSKSKMRGFSKXXXXXXXXXXXXXXXXXXXX 80

G++K L + VL +DIITG+ KA K L S+ + G+ +K

Sbjct: 17 GEVKYLDLMLVNIIDIITGVKAWKFKELRSRSANFGYVRKMSFLVVIVANAIDTIMD 76

Query: 81 XKGLLMITIFYIANEGLSIVENCAEMDVLVPEQIKDKLRVIKND----TEKSD 131

G L T+ +YIANEGLSI EN A++ V +P I D+L VI++D TEK D

Sbjct: 77 LNVLTFTATVLFYIANEGLSITENLAQIGVKIPAVITDRLHVIESDNDQKTEKDD 131

>gi|141088|sp|P26835|YNGD_CLOPE HYPOTHETICAL 14.9 KD PROTEIN IN NAGH
 3' REGION (ORF2) >gi|1075967|pir||S43905 hypothetical
 protein D - Clostridium perfringens >gi|455154 (M81878)

311

ORF D (Clostridium perfringens)
Length = 132

Score = 60.9 bits (145), Expect = 4e-09
Identities = 38/127 (29%), Positives = 63/127 (48%), Gaps = 3/127 (2%)

Query: 1 MNEVKFRFTDSEAFHFIY-AGDLKLLYFLFVLMFVDIITGISKAIKNNLWSKKSMRGF 59
+N +K+ +I+ A D+ L+ L V +F+D +TG+ K K+ L S +RG
Sbjct: 5 INYIKWGIIVSLGTLFTWIFGAWDIPLITLL-VFIFLDYLTGVIKGCKSKELCSNIGLRGI 63
Query: 60 SKKXXXXXXXXXXXXXXXXXXXXKGGLLMITI-FYYIANEGLSIVENCAEMDVLVPEQIKD 118
+KK + I ++YI NEG+SI+ENCA + V +PE++K
Sbjct: 64 TKKGLILVVLVAVMLDRLLDNGTWMFRTLIAYFYIMNEGISILENCAALGVPIPEKLKQ 123
Query: 119 KLRVIKN 125
L+ + N
Sbjct: 124 ALKQLNN 130

>gi|2293160 (AF008220) YtkC [Bacillus subtilis]
>gi|2635548|emb|CAB15042| (Z99119) similar to autolytic
amidase [Bacillus subtilis]
Length = 134

Score = 36.4 bits (82), Expect = 0.099
Identities = 25/109 (22%), Positives = 41/109 (36%)

Query: 17 FIYAGDLKLLYFLFVLMFVDIITGISKAIKNNLWSKKSMRGFSKXXXXXXXXXXXXXXXXX 76
P + G L LM ++ I+ K + L KK KK
Sbjct: 20 FFFGGFQYSFLILLSLMAIEFISTTLKETIIHKLSPKKVFARLVKQLVTLALISVCHFFD 79
Query: 77 XXXXKGGLLMITIFYIANEGLSIVENCAEMDVLVPEQIKDKLRVIKN 125
+G + + I +YI E + IV + + + VP+ + D L +KN
Sbjct: 80 QLLNTQGSIRDLAIMFYILYESVQIVVTASSLGIPVPQMLVDLLETIGN 128

>gi|1181973|emb|CAA87743.1| (Z47794) holin protein [Bacteriophage
CP-1]
Length = 134

Score = 31.3 bits (69), Expect = 3.3
Identities = 27/88 (30%), Positives = 36/88 (40%), Gaps = 5/88 (5%)

Query: 29 LFVLMFVDIITGISKAIKNNLWSKKSMRGFSKXXXXXXXXXXXXXXXXXXXXX--GGLL 86
LF L+ D ITG KA K S ++G K G +L
Sbjct: 18 LFALILPDPFITGFLKANKWKVTDSTGLKGVIKHTLTFFIFYPVAVFLTYIHAMAVGQIL 77
Query: 87 MITIFYIANEGLSIVENCAEMDVLVPE 114
++ I Y A LSI+EN A M V +P+
Sbjct: 78 LVIINLYYA---LSIMENLAVMGVFIPK 102

Table 21

Phage 182 complete genome sequence. 17833 nucleotides.

```

1      tagaattattg tcataaaaca caaacataat aatgcatatt attgtttaca aatatgtaat ttcgtgatat
71     aatataatttg taagttaaag gaggtgacaa aagaacaaat cataaatgct tttagaaattg caaaaactat
141    tggaggaaaaa ataatagaaat attcactaca acaaatagat gaaatttaaat caacaatttt cagaattaga
211    ttaaaaaggcg atgaactaga ggaattgggtg gacgaagtaa acgatattgc taaagatccg gaggaagat
281    atctttttatc gttttattac acagaagaag aacgtttgtt tgaaattccc tctgcaagat taatagatta
351    ttacaaacgaa aagatcacaa atctgaaatc ggaatcata tcaactcgaa aaagattaca aaaacttagta
421    aaataattac acaaaaagct ttacaaatat aacacatcat gttatactaa aagagtagta agggaaacgga
491    aaatacctta cttcacacct caatcattct tatcaaaata caaaaggagg gaaataaatg ggtcgaaaaa
561    taatgcacacg aaacgtaaca tcaactaaag tagaattctc agaagttatc gtacaagatg gagcgccaac
631    aattgtacca tgcgaaccag ttgtcttaac aggaaaactt tcagaagaaa aagctttatc agcgatcaaa
701    cgtaaaaaac ctgataaaaa cgtagtgtga acaaatgttt cacatgaaac agcgctttac acaatgccag
771    tcgataaaatt tctcgagtta gcagacaaat caacacaaag ctaataaaaa caaaactaaa acaaaaacaga
841    ggagattata atcatggaaa tctgtaaaaag cacatttgac acacaaacac cagaaggaaat gttacaagta
911    tttcaatgcc aaaaacggggc tttcaattccg ttacgtaacg caattggcga agtactagaa ttgaaagata
981    tttctagttta ctcagacgaa gtttctgggt ttggtggagc cgaacctaca caagcagaac tagtcgcttt
1051   cttcacagaa gatggtaaaa cttatgcggg tgtatcagca gtacgaacaa aatcagctaa aaacctaat
1121   gatattgata ctgctaacc cgcacataaa ccaaaaattt cttttgtcga aggaaaaatc aacggtggac
1191   aaaaattttgt aaatctacaa gtggtttcac tgtagcataa aaatacagga atctagtaag ccaatagcgc
1261   aatctcgcta ggtggttttt attatgtttc tacattgagg tgtgtagaat tgaccgtaag aatctcaag
1331   aatgatagag ccaagttaga gaaaatctac ggttaaatct acaaaagctc taaaaaaaac aatcgtttaa
1401   gacaaaaagg agttgaggaa aggcaacttc caactgttcc aacatcaag aaaagacctt ttgactacgt
1471   aaaaacaaca aatatgagtc gtagtgattt taacaagatg tttagcaggt tggtagattt tgcacaacct
1541   tacaacgaga attacattt ttgagatcaac aagcgaatg ttgcaatctc aagagcgcaa atcaaaagaag
1611   cgcaaattaa aacagagcaa gctcaaaaag cgaagaaga acactacaa gagcttaaca aagttgaagt
1681   taagaagccc acagaaaaa caattgtcac accaactatt ttaacagagt taggtgctga cttacctttt
1751   caagcaatat cagattttta ttatgcgctc ttcaacttcc cagaaggagt tcaagtctat ttgaaaaata
1821   taggaaaaa agacgaacaa tattttgacg aaagagacca actttattac gacaatttca gacaagcgat
1891   gtttactatt ttcaattcag acgctgacga tattgttcgt ttacttgact caatggggct tgaatctatt
1961   atgaaaaacat atgttagtaa cttcttagac atgaaccttg actacattta tgacgaagca gaagtacaac
2031   agaaaaaaga acaagtttac agtaagattg caaaagtgat cgagtctgaa acaggtggag aagtcacctc
2101   atataacccc acgaagaaca tcacaattaa ttcagaaaaa ggagaaagat tatgattaa aatatactgt
2171   gcgactttga aacaacaact gatctcaacg attgtcgtgt atggtcgtgg ggcgtatgca atatagacaa
2241   cgttgacaat atgacgttcg gtttagaatt cgattctttt tttagtggtg taaaatgca aggcagcaca
2311   gacatttatt tccacaacga aaaaattgac ggagagttta tgctttcatg gttattcaaa aatgggttca
2381   aatgggtgtaa agaaacaaaa gaagatcgaa cattctccac actcatatca aatatgggtc aatgggtatg
2451   tttggaattt tgttgggaag ttaattacac aacaacaaaa tcaggtaaaa cgaaaaaaga gaaatctcga
2521   acaataattt atgatagcct taaaaaatat ccttttccag tgaacaaat tgcaagaagt ttaattttc
2591   ctataaaaaa aggcgaataa gattatacaa aagaagagac tattggttac aaaccaacaa aagatgaatg
2661   ggagatttta aagaacgaca ttcagattat ggcgatggca ttaaaaattc aattcgatca aggactcaat
2731   cgaatgacta gaggaagcga cgtcttaggc gattacaaag attgggtaaa agctacacat ggaaaatcaa
2801   ctttcaaaa atggtttctc attttgtctt tagggtttga taaagactta cgtaaaagcat aatgggtatg
2871   cttcacttgg gtaaacaaa tttttcaagg gaaagaaata ggtgacggca ttgtcttga tctcaactct
2941   ttgtatccct ctcaaatgta cgttaagacct ttaccatag gaacacctct attctacgaa ggagaataca
3011   aaccgaacaa cgcactacgc ctgtacattc aaaaatcaa agtaagattc cgtttaaagg aggggtatat
3081   tccaacacat caagttaagc aaagtctact attcatcaa aacgaatc ttgaatcaag tgaatacaag
3151   ttaggagttg acgaattaat cgtcttact cttaaaaatg ttgacctaga attattttt gaacactacg
3221   atattttaga gatacattac acttacggat atatgttcaa agcttcttgt gatattgtca aaggctggat
3291   cgataaattg atcgaagtaa agaaccacac cgaaggggct agaaaagcta acgccaagg tatggttaaat
3361   agcttgatag gaaagtctcg aacaaacctc gacattacag gaaaagtgcc ttacatgggc gaggacggca
3431   ttgttcgatt gacactagga gaagaagaat taagagatcc tgtttatggt ccgcttgcta gttttgtgac
3501   ggcttggggg agatatacta ccattacaac cgtccaaaaa tgttttgatc gcattattta ttgtgatata
3571   gatagcattc atctagtagg aacagaagtt ccagaagcaa tgcactactt ggttgatcct aaaaacttg
3641   gttattgggg gcataaagc acattttcaac gagcaaaat cattcggcag aaaacatacg tagaagaat
3711   tgatggcgaa ttaaatgtaa agtggtctgg tatgcccag cgaataaaa agattgtaac ttttgacaat
3781   tttgaagttg gtttttcaag ctatggaaag ttgctaccta aaagaacaca aggtggcggt gtattagtag
3851   acacaatggt tacaatcaaa taaggaggac taataatgga actatataaa gcaatgttta tctgactgta
3921   tgagggtact attgacggtt acgatactga acactatgta gatattttt tacatgactt tgaagaaata
3991   tctggaaaaa aaacacgtga aattgaagca gtaacattag taaaaacagg aaatttataa aatacaattc
4061   tttacatcct ttgcaaaagta tggtaaaaata tttctgtgat agtgacaag agtcaaat tggcgagattg
4131   ggcgaaatga cacgtgaaat atcgtgcgct cccgttaagt tatggacaca taaacgtttt gaccgtcaac
4201   caatcgcaaa aaccttttag gtagtgcctc taaatgtggc tactctttt ttggtttcac agattatgt
4271   ttcacgtgaa acagttttta tggataataa gaatcaaaag gaggtggaga ttatggaat taaagaacat
4341   gaatcaattt taaatgggat tcttgaagag gtcacagacg gtgaagcaag atcaaaagatt gtagaacatc
4411   ttgaagcatt cgcgaagacg tacggagcaa caactgaaag tttgacatca gcaaatagca cacttgaaaa
4481   gttcaagaaa gataacgaag cgttgggttat ttcaaactca aaattgttcc gagaacgagc gatcgtagaa
4551   ccagcagaaa ataacgaacc agaaaacagc cagaatatta cactagacga tttaggaatt taaggaggaa
4621   aaaaatgggc tgacaaaatc acagaacaa atgtttcttcg tgccacaaat gtagaacac cagtacaatt
4691   aatgactgct atttataata gttcatcatc tctttttcag gcgaacgtac ctatgcaaaa tgcagataac

```


4761 atcgaagcgg ctgggtgcagg gatcacacgt tttagacgtag taaaaaacga atttatttca accttagttg
 4831 accggtattgg taaagttagtt atccgataca aatcttggcg taaccttttg aaaatgttta aaaaaggaaa
 4901 catgaccttta ggtcgaacga ttgaagaat tttgtgtgac artgcacagg aacataagtt caacctcgac
 4971 gagtctgtta caggggtart taaacaggaa gttcccgatg taaaaacatt grtccacgaa attaatcgtg
 5041 aaggttacta caaacaaacg atccaagaag catgggttaga aaaagcattt acctcatggg ataatttcaa
 5111 tagtttcgtt gctgggttaa tgaacgcttt atacacaggt gacgaagtaa gcgaatttga atacacgaaa
 5181 ttattaatag caaacaccca agaaaaagag ctattcgaag agatcgaaat tggcgaattt actgaatcaa
 5251 atgcataaaga atttatccgt aagatcaaat caacctctaa caaattagaa tttatgagtt cgccttaca
 5321 cgcctcaagga gttaaaacat ctacctcaaa atctgatcaa tacgttatta ttgacgcgga cacagacgga
 5391 accattgacg ttgacgtttt agcagcggca ttcaatatga gtaaaactga cttttagtaga ccagaatcgc
 5461 ttattgatga gtttccctaaa aaagaaggcg aagaatcgtc aatatctgtg gcagttattg tagatagtga
 5531 atgggtttatg atctacgaca aattgtacaa aacaacaagt ctatacaacc ctgaagggtt atattggaaat
 5601 tatgtgttgc accaccacca actatatctt acttctcaat tcgggaacgc tgttgccttt gttaaatcag
 5671 caacaaaacc tgtcacaaaa gttgcttttg caagtgcac aactagtgtt gttaaaggat catctaaaga
 5741 tatcgcttgg acattttacac cagttagaagc aacaaccaa caaggagaag ttgtttctac agcaccagca
 5811 ttgggttaagg caaccgtaaa acaaacagca ggtaaagcga ctgcccgaac cgtagaaggc ttagaagtgc
 5881 gtcaatcatt agtaacattc acagctatcg gaggctcaaa agcaacggtt cttgttacgg ttactctga
 5951 ctaaggagga caattatggc aagaaggtat acaaatgtaa aattgttggc taactgtgct ttgataaca
 6021 cctatacaca cacaagatgg tttaaaactc aacaggaaca ggaatcgtac tttaatctgt ttctgttct
 6091 taacaggaat agagattggt ctatatcaag ggatcacaaa ctggtgggag tttttagagt agataaacac
 6161 aaagacgcct tatatgcttg gaatataga atgacaacac aagtttctgt acctttgaaa ttgatgtttt
 6231 atgctcttgg tactgatatt gaatataga atgacaacac aagtttctgt acctttgaaa ttgatgtttt
 6301 acaaacctat cgtttcgtata ttggtatcac agaaagtctt attgcaaaag aacaccttca acctttat
 6371 tcgaattgga tacctttcat taatacaatt gaagagtcgc ttgattacgg tagagaatca acaacaacaa
 6441 atgtaacaac ttttcatcct aacgatggag tcaattttct tgttattcta acaagtgaag caatgccagt
 6511 tggagataag gaagataaat cagggaggtc aatagttagt ggcctatctc ctttttctta ttatttactt
 6581 cctatacaat caagtgggga ggtatcaaaa ccaaatgggg caggcaatgc taattttgga ggtacatgg
 6651 cgtttcttac aacgaagaa ccttttctaa ataagatagt cgggatgtat gtaacgtcgt atacagggtat
 6721 accattcatt tgggatcacg cgaacaaaac ggttaggtat aatgcaggag gttcttataa gatcagcgtt
 6791 ccaacctacg ctagtatcgc aacaggaaca atgaaaacat tcgctttctt ttgtgtaaaa gaagcaagaa
 6861 cattcgtacc taaagaactt gatctttagt ggaacgtgta taactacttt agagaagcgt ttccgtttaa
 6931 tgttaaggaa tcaaaaactat ttatgtatcc ctattgttta atagaatata cagatacaaa aggacatgta
 7001 atgactttaa gacctgaata tcttaccagt ggttaattga gtgtatatgt aaaaggttctg ttaggaaatt
 7071 ctaataaagt gatgatcgag ccgattgatt atgatgtaag taactcaacc attattacca atttaagtga
 7141 caagatgtta atcgataatg atcctaacga tgtaggaggt aactctgact atgcttctgc attcatgcaa
 7211 ggaacacaaa actccttgat tgcctcaagag caaaaacattt gcaatacttt cagacatggg atggcaaaa
 7281 gtccaatgag tacaggagga ggcgatcttt cagccttagc aagtaacaa cctttgtgtg gtttgcataa
 7351 catcatggga cgaggagaca aagtaaacaa ctatgtttct gaaaaagaaa acggtttgaa cctctggga
 7421 ggttaagtgg cagatatoga aatatctcca gataatgtaa cacagcttgg atcaaaacta tctttcaca
 7491 caggaaactt tcaaaactat tatcaattgc gcttcaaaac aattaaatag gagtatgcaa caagacttga
 7561 tcgttacttc tcaatgtatg gcacaaagag caatcgagta gctacaccaa acttacaacac aagaaaagca
 7631 tggaaatttca ttaattttaa agaaccacaa attgtaggca caatgagtaa cgtatgtata acacgtgtga
 7701 aacaaatttt tagtgcaagg gttacgcttt ggcatacga tgaatgtttg aattataacc aagacacagg
 7771 agatgtatag gaaggaggaa taagatgagt agacgaagaa gtcaggagct tgctagaat aaccgttata
 7841 cagcaaaaag cagaccttat ccaaatgaac cctattcaag tgaatgtaga gaatcagctt actatgaaca
 7911 ttatcgtaga caactcacgc tccctacggt tcagttgttt gaattgggaaa atttgccaaa atcaattgac
 7981 cctcgttatt cagaatttat ttacacactt aatgggtatc ttggtttctt taaagacctt acacttgggt
 8051 tcatggtttg cgcaggggca gaagatgggt aaatcgatca ttatcacaa cctattttct ttacagcaaa
 8121 cgaagcaatg tatcacaga gatattcctgt ttttaagatat gatgatgatg atgataaatc aaaaatgtatc
 8191 atgtgtgata ataattgact gaaagtccct acgttaccaa gtttcatatg ttttgcctta gatattggcg
 8261 acataaacca gatatacga gtgaatcgaa gagcgcaaaa aacacctgta attattcaca ctgatgaaa
 8331 gaaatacttc tcatgtctac aagcttataa ccaaatgac gaaataatc aggtctgttt tgtggataaa
 8401 gatattggag ttgacgaatc ttttaactga tggcaaacaa atgctccata tgaatagat aaactacgat
 8471 cagaattgaa cgaagtatgg aatgaagtgt taacttttct aggtatcaac aatgctaacy tagataagac
 8541 tgcacgtgta caaacatcag aagtccttct taacaatgaa cagattgaaa gttcaggtaa catcttctga
 8611 aaatcaagaa aagagttttg cgtatcgtga aatcgtgtct ttggcgatga acttgacgga aagattgacg
 8681 tgaagttag aacagacgct gttcgacaat tacaactggc ggcagggtcaa tcaaaaaaag accagatgag
 8751 tggagggttg ccaagtgcct cttaaacggt atattgaaag ttctacttat taccaacctg aattatctcg
 8821 aaaaagacgt attgaagttg gccgaaaaca attgtttgat ttgtattatc cgttttatga cgaacaaaa
 8891 cgagcagaat ttgaacaaaa atttatcaat cacttttact tgagagagat aggtctcaga acgatgggat
 8961 cattttaagt taactctgac gaatatattaa atctaaacat gccctattgg aataaaatgt tccatcaaa
 9031 tcttgaagag tttccgattt ttgatgacat ggactacacc attgatgaga aacagaatgt gttaaatgag
 9101 attgatataa acatcaagc gaatcgtgat gaatcgaaga accaaacgaa gcaagtagat caaacagaca
 9171 acagaacaaa aaatacacgt gacacaggaa caacgatctc tttctcaagg aacacttata cagacacccc
 9241 tcaaaaagat ttgagaattg ccagcaatgg agatggaaac ggtgttaatca attatgcaac aaatatcaca
 9311 gaagatttga gtaaaagaac aacaagctcc acaggcggtg aacaaacaa cgacaaaaca aatcaaaata
 9381 cacgaagcaa tgcctctgaa aaagaacaaa agaacacaga cattaataaa gatcaaaatc aaccaaaga
 9451 tacgattaca cgaataaag gtaaaaagg aaacactgat tatgctgact tactcgaaaa atacgttaga
 9521 agtgttttga gaattgagaa aatgatcttt agagaaatga acaaggaaag cttatttctc cttgtttatg
 9591 gagggaggta gcaacaatgg tagattttaa ccccgacaag cgttttagc gtttaccgcg ttgtattcaa
 9661 gaacgcttta gcaaatatcc tcatatgaa tacagatag aattactatt agatgaagaa gtatcggtt
 9731 taattgcta tctgaatgaa gttgtgtctt tagttaatga tatgagtggt tatttaaat accttatcga
 9801 acattttgtt gagaagttag aagagatcac aaatgacaca ctcaaaaaat gggtgtctga tggtagctta
 9871 gaaaatttaa tcaatgatag ttttttttga aattatatca aagaaatcaa aagattacaa atcttgggtg
 9941 ctgaaacacg tgcataacgt gtgaatattc ttttgacaaa aaataaacgg gatgttctg atgacgaac
 10011 attttgggtat aagattcaac gcgacaatac tgattatgga gccgatccta ttgacacgtt acgtattgtt

10081 gcaatcaata aagtttagtgg ctggaatacc gctacaggag atattttatct taacattaaa ggaacggagg
10151 gtgtataatg gcagacatta gaacacaact aacaagtga gatggatcag acaatttatt tccaatttca
10221 aaagccgtta atattatgac taatagcggc acgaatgtag aaggagaatt gggtagactc aaacaaaatg
10291 acgaaacaat gaataacctca gtccaaaatg ctgtagttac tgccaatcaa gcaaaagatt ctgtagctga
10361 attaaatgta aatgtttggtta aactaaccac tcgaataaca acattagaga gtacagtggc taactctgat
10431 ggtattcgct atgtagaggt gtaatatggc agataaaaat attcaaatgc aggataaaga tcataatcgt
10501 ttaatgctcg ttacaattgc taaaaatgct ctaacaggcg actctaactc tgaattagct aatgctgaaa
10571 taagaggtta cgctagtga gctaaaacac ttgcacaaca agctaaagaa actgctgctg gtttgcacac
10641 agaaattgac acagtacat caaccgcgaa tcaagcgttg acgaaggctg gtacagcaca acaaacgcga
10711 gaacaagcga aaacaacagc aaacagatc agcgcagttg caacggcagc taaaaacaca gctgattcag
10781 cacaaaaaag tgcaactgat ctactgttcc gagtaagcag tttagaggac acagcaatc aataactgt
10851 attaccatag gaggaaaaat aatggcaaat aaaaattatc aaatgaagga tagcaatgac aataattat
10921 atccaagtgt tcgagcagaa aacttggttag atttgaccag tcgtgctgaa ttaacaatga caaattgtca
10991 attatctgca gctggtgata aaacaattgc aatctcttat ctggtgtcag taggtatgct gtttgcacac
11061 ataaagtta ctgaaagtct gacaaacctc gtgatcaca cgctaccaga aggttttaga ccaataagaa
11131 caaaacgtat tgggtgttcc gcaaaatatt acacaccaa tccaacagat acaaaagaaa tgggtttatgt
11201 atcaatcaca cctgatggca aagtaactgt aaatgacaat gtaggtaaaa tcgaatatct atccctagat
11271 aatgctgctt tccctctaaa ataaggaggt tcataatgaa gaacgaattg atattcaaat gaacaagatg
11341 aaagaagaaa atcaaaagaa ttacctatg caccctgaaa cgaacccgaa acaagttgt ttgtatgaaa
11411 cattgcatgg aaatgaaaat caggagaggt tcaacaattt tgttgacaca agaaaaatga caactacaat
11481 tgatgtaagt gcttatgggg ttatcgctga cgtgttaaca gattgtacac caatatcaaa taacttactt
11551 gaagaaaaaa gcgaaatggg tatcactttt tatttctctc cttgtgaacg tgattcatat ttgtgctttg
11621 ctaacacccat tgaattgaaa cgtgatgtac ctgtagttac ttcttagga tcgggagaaa cgacattaaa
11691 gtttgaacaa atgacggcat ttaagttaaa catcgaaagt ttcaatatgt atgggtttgc attatgggtg
11761 ccacaaggcg ctcaaatggt taaaggaaat ttctttaatg atactcgcaa ttacaactcg ttgactttg
11831 atttgtttgt tcgttaactgt actttaaatg aaggaaactg tggttgtgtt gctagaggta gaggggttac
11901 atttgaaaaat tgcctattct ctaatatctc tcaagcaatt atcaaaacag cttttcccg taataatggt
11971 atgtggcaag ggaacgatat caactactgg ggtacaggtt tttagagttt ctttgtgaaa acaacccgta
12041 ttcatttttg tacagcgatc attatcgaca atgacgatga ttatcagaat gtaattaat tctgtgaaat
12111 ttctggtaac acaatcgaag gtggcgtaag ttattatcga ggatatgcgc ataatgtca tgcctaaac
12181 acaaacatt ttctagcata cggaaataga aaagctttgt ttgagtttca agatgtggat caagcttata
12251 ttgatgtaga tgtttattgt cgttaactac aagtcgaggg aatgaatagt acagctattt cacgttcaat
12321 tggttgttac ggacattacc gaaacttaaa gattacaggt aaattatctc gttgtcaagg acatgtatc
12391 acgttgtatg gcggtggcgt taattttctat tgtgacttga tggcacaaga agcacctttg acggacggtt
12461 accggtttat tcaaacggct gacaatcgag ttaactatga tgggtttgtt gttcgtgggt tgcctaattc
12531 aacaaaagta aatacaccac tgatctataa agcacctcag actgtttctt ataactgtag aatcgatcat
12601 gtgctaacag gtccaaatgc aagtaatgta tataactagg aggatatgag atggcaactc ttacaatga
12671 acaaatagct agaggacaaa caatcgctaa aatactttca aaatatggct atcaaaaaa ttcaacagta
12741 ggagttgtcg ccaatctcca ttgggaatcg gctggtttga acccgaaacg caatgaatat ggtggaggcg
12811 gatattgggt aggtcaatgg acgcttaaaa gcaatcttta tcgccaagca caaatttgtg ggtgtctaa
12881 tgctaaagct gaaacgttgg aaggtcaagc agagatcctc gctcaagggg ataaaaacag tcaatggatg
12951 gataatacac ctgtttcttc tgcaggttat actaacctc agaccccttc agcatttaaa caactcgaa
13021 atattgatgt tgcataactt aattttatgt gtactggga acgccccgtt aaactcata tcgaagaaag
13091 acctgatctt gcacaagctt atagtaagca tattgacggt agcgggtggc gtggcgtaaa acgttgcctat
13161 ggaaccccaa tcaagaatac aaatcgttat cctaaaagtt tcatgagtgg acaacttttt ggacgcgatg
13231 caggaaaacgg cagaccaaat aatttccatg atgggtttga ctttgggtca attgacacc ctggcaatga
13301 aatgattgca tgtttcgatg gaacagtaac acatgttggg acaatggggc cattaaagagc gatatttgtg
13371 ataaatgatg gtacttaaca tatcggttat caagaattta gttataacca gtcataataa gttatgttaa
13441 ttggcgacaa agttaagaac ggcaaatgtt gcgcaatcag tgacggcgat catttacatt taggtttcac
13511 taaaaaagat ttattgactg cgttaggacc ttctttcata gatgatggaa catgggaaga ccttttgaag
13581 tttttagggc aatgttttgg agatggagat actggcgagg ataatgacga taacaataag gataaaaatg
13651 atcttattta tctattgcta tccgatgcct tgaatgggtg gaaattttaa taaggagaaa aaggtatgat
13721 agaatatc acacaatggg tggcagatga taactcatct gtttatgggt tgatttatg tttatgggtt
13791 gcaatgatta tgcattttgt gttaggtttt acaattgcca aatttaacaa ggaaatcgac tttagtagtt
13861 ttaagctaa agcaggtatc attgttaagg tggcagaaat ggttttagtg gtttacttta ttctgtatg
13931 agtaaaaattc ggtgcagtag gtattacaat gtatataaca atgttgggtg gtttgatttt atcgaattt
14001 tatagtatac taggacatat ttcagatata gatgatgata ataattggac tgattatggt aagaagttt
14071 tagacggaac actcaacaga aaggacgata ttaaatgatg aatgggtatg atatctctag ttatcaaaa
14141 ggaattgata tttcaaaagt tccatgcgat tttgtaata ttaagcaac aggcggaaca aggtatgtaa
14211 accctgattg tgaccgagca tttcaaacag ctttgccttt aggtaaaaag attggtgtgt atcattttgc
14281 gcatgagagg ggtttagaag gtacacctca acaagaagcg caattctttt tagataatat taagggttac
14351 attggtaaaag ctgttcttat tcttgacttt gaagggtcaa atcagaaga tgtaaattgg gcgaagcat
14421 ttcttgatta tgtttataat aaaacaggcg ttaagcatg gttttatagc tatacagcaa acctcaatc
14491 aactgatttt tctagttatg caaaaggcga ttatggttta tgggttgctg aatatggatc aaatcaacca
14561 caaggctact ctcaaccagc gccacctaaa acaataattt ttccaattgt tgectgtttt cagtttcaaa
14631 gtaaggagcg tttaccagga tacaacggca atcttgattt gaatgttttc tatggcgatg gtaatacatg
14701 ggaatctgat gtaggtaaaa aacaggatca aattgttctt cctgaaaata aaatatttga cgccacaagt
14771 gatgagttta ttttcaactc tacaacaggt agcacagcg tgttttattt tgacggagaa acgatctttg
14841 aattgtctga tccaacacaa ctgcatacata tttagggaaac atacaatcat gttcatggaa aagaattccc
14911 atcaatgggt tggacacctg acaaatttga tatttctta aaatgtatg aaaaagaaacc agtatataa
14981 taggagtgta tagtatgaca aatagcttag gcgttaaaat tgaagagaaa aacttactat ataaccctaa
15051 caatgcttta ggttttaatt gcttaaatgt gtttgaata ggcgcacgtg gtataggtaa aacttatggt
15121 tataaaaaat ttgttgttaa tgcgtttatt aaacacggcg aacaatttat ttatttaaga agattcaaaa
15191 cagaacttaa aaagattcct caatttttca aaacaatggc gaaagaattt cctgatcata aacttgaagt
15261 aaaaagaaaa gaattctatt gtgatgataa attaatgggt tgggtgtgtc cacttagtac gttgggaatt
15331 gaaaaatcta atgaatatcc cgaagtctgt acaattttgt ttgatgagtt ttaatttgag aaatcaaaaa

15401 tcacttattt accaaacgaa gctgaagcct tattgaacat gatggaaaag gttttccgaa gacgtacaaa
15471 tacaagatgt gttatggtga gtaatgcaac tagtgtagtg aacccttatt tcttgatatt caatctgcag
15541 ccagatttga ataagcgttt taatctatat caagatcgag gtatatgat tgaattgrgt gattcaaaaag
15611 actttgcaga agtgaagaga gaaacacctt ttggtagatt gattcgtgga acagaatagc aagatttttag
15681 tatcaacaat gagtttgta atgatatgta tacgtttatt gaaaagagaa gtaaaaatag tagtttctta
15751 tgcgccattg ctttcgaagg gaaaatcttt gggatttggg tagacgctga aacaggttgt gtctatgtga
15821 gttatgatta tcaaccaaact acaaatcatt tttatgcaat gactacgaaa gaccatgaag aaaatagatt
15891 gctgatgaaa aattggcgaa ataattatta tctttcaaca gtggcgaaaag cattcaagaa tagttatctg
15961 cgggttgata acattgttat taagaattta cattatgatt tgtttaataa gatgaaaatc tggtaacctt
16031 attttagtag agctaccacg attagtctta ttacaatgat gaatagtaga taacatagta attgtagtct
16101 gcgatatgtt tgttttgggt ctttggcggt agtgattttt gctaagcctt ttttgtttgc ttttggatcg
16171 ggtgtgttaa tgtagacgaa atcttttctc atagtctctt ctcccttatac agttttaata attcctgtta
16241 aaatgtagct ataggacgtc catttcttct tattctaacg caattcacta tatccatttc taggtatata
16311 cggctatatt ttaatgcttt tgttaagggt agagggtcgg ttttgtgtat caaaacctcc caacctatcta
16381 tataaaatag tgtgatctg tatattgggt ccttgtagaa tgtagccatt attccacctc ctttaaatag
16451 ccttttggta tttgtaacgc taactgtag cgagaacca cttttacgta tgaagtact aatttcattg
16521 cctgacaata cttttcaaga atgttaaat gactcgattc gggtaaatgc gttgaatgag ttaacaaaag
16591 ttcgggtgata tttatttccg gaacgtcgaa atcttgtaaa gtccctctta tgaatctctat ttttccattg
16661 tctgaaaggt tacgtttaca gtagaacgt aaccattcaa ttagttcgcg gtgttctttg aatgttcgtg
16731 caatcatttt aattcctcct atttgcctgt aatttgttta tatccgtcat gtttcaattg ttccgcatag
16801 tgttcaacgc ttttcattga tttcgttatt gcgatatata tgcaatggct atcaagataa acatagttat
16871 atttatcatg tgttaacacg aactcttttg taacgtaatc aatgtataaa attaatgtt ttctctcttg
16941 tgttatttct gacttgatag acgctaaaact atcgttgta cttttagttg gttgatttaa accctctaaa
17011 attaatgata aattgttaat catgtaaaac actcctttta tattaatttg atattgatac caccaatcga
17081 ataagatttg tagcattgta tgaattta atgttatttc tgtagtttc catgaatact cggaaataag
17151 atccatatct aattccttta gttcttcaaa agataacaaa caatattcct catcgccctac ctcatcaata
17221 tcaataagat aatgtttatt gttttcggtt cctacgatat gataattcat atcccactca ttaagggggt
17291 gaagttagaga tacctctcct ttttcagcta ttaatgattt attgttcata tgaacacact cttttatatt
17361 aatttgatag tgataccacc aatcaaatgt gattggtagc attgtattaa attaatattc tggataattt
17431 attgagaaaag tccagttatc atcaaatgaa attgttttat tttcaagtaa ctttttagcc tcatccacct
17501 caaattctaa atagaggaa ttactaagtt tatcctcatc tctaaaaatt ttcatacata ccacgttatt
17571 tgaataaatt tctgtgtata cgatcggttc attcatgttt atcatccttt ctttattaca tataatgat
17641 atcatgtatt cacatatatg tcaatcattt aattcattta ttttaattgat ttatttgatt gtttttttat
17711 gatcctttct ttattacatc tatattatat catgtatgat tgtatttgc aacaattaaa ttcatataaa
17781 tgtagtttgg ggtcagttac atttgtgtta tcaaaaaag ataatttct att

Table 22

Phage 182 ORFs list

nb	Name	Frame	Position	Size (a.a.)	Key words
1	182ORF001	2	5966..7780	604	Tail protein;
2	182ORF002	1	2152..3873	573	DNA polymerase;
3	182ORF003	1	11305..12639	444	
4	182ORF004	3	4626..5954	442	Major head protein;
5	182ORF005	3	12651..13700	349	Glycyl-Glycine endopeptidase; Lysostaphin precursor;
6	182ORF006	1	14995..16026	343	Encapsidation protein; ATG/GTP-binding site motif A;
7	182ORF007	1	7795..8775	326	Upper collar protein;
8	182ORF008	2	14105..14983	292	Lysozyme; Muramidase;
9	182ORF010	2	1310..2155	281	Terminal protein;
10	182ORF009	2	8765..9601	278	Lower collar protein;
11	182ORF011	1	9607..10158	183	Pre-neck appendage protein;
12	182ORF012	3	10872..11294	140	
13	182ORF013	1	10456..10860	134	
14	182ORF014	3	13716..14108	130	Lysis protein;
15	182ORF015	2	854..1225	123	Early protein;
16	182ORF018	-2	18429..16737	102	
17	182ORF020	3	10158..10454	98	Leucine-zipper motif;
18	182ORF019	3	4323..4613	96	Head protein;
19	182ORF016	-3	16749..17033	94	
20	182ORF022	1	12868..13149	93	
21	182ORF023	-2	11914..12189	91	
22	182ORF017	1	154..426	90	
23	182ORF024	3	6174..6446	90	
24	182ORF025	2	548..814	88	Early protein;
25	182ORF026	-3	12999..13259	86	
26	182ORF027	-1	14642..14896	84	
27	182ORF028	3	14430..14672	80	
28	182ORF021	-3	17106..17339	77	
29	182ORF030	-1	16199..16429	76	
30	182ORF031	-3	8379..8603	74	
31	182ORF032	-1	11195..11413	72	
32	182ORF033	-1	4727..4942	71	
33	182ORF034	-1	5951..6160	69	
34	182ORF029	-3	17412..17606	64	
35	182ORF035	-3	15570..15758	62	
36	182ORF036	-3	2127..2315	62	
37	182ORF037	-1	12095..12280	61	
38	182ORF038	3	14769..14951	60	
39	182ORF039	2	9992..10171	59	
40	182ORF040	-3	16029..16202	57	
41	182ORF041	1	3886..4056	56	Early protein;
42	182ORF042	-3	10671..10832	53	
43	182ORF043	-3	10491..10652	53	
44	182ORF044	-1	6299..6457	52	
45	182ORF045	-2	6571..6729	52	
46	182ORF046	2	2372..2527	51	
47	182ORF047	-2	13201..13353	50	
48	182ORF048	-3	3243..3395	50	
49	182ORF049	3	1578..1724	48	
50	182ORF050	2	8012..8155	47	
51	182ORF051	3	9390..9530	46	
52	182ORF052	1	4096..4233	45	
53	182ORF053	2	15656..15793	45	
54	182ORF054	-2	8002..8136	44	
55	182ORF055	2	8324..8455	43	
56	182ORF056	3	6549..6680	43	
57	182ORF057	-3	8133..8264	43	
58	182ORF058	-1	5048..5176	42	
59	182ORF059	-2	15748..15876	42	
60	182ORF060	-3	15276..15404	42	
61	182ORF061	-3	1974..2102	42	
62	182ORF062	-2	1867..1992	41	
63	182ORF063	-3	14181..14306	41	
64	182ORF064	-2	7234..7356	40	

317

65	182ORF065	-2	3460..3582	40
66	182ORF066	1	4234..4353	39
67	182ORF067	-1	13763..13882	39
68	182ORF068	-1	7148..7267	39
69	182ORF069	-3	4908..5027	39
70	182ORF070	-3	912..1031	39
71	182ORF071	2	11741..11857	38
72	182ORF072	-3	11610..11723	37
73	182ORF073	-3	2763..2876	37
74	182ORF074	-1	8813..8923	36
75	182ORF075	-3	7353..7463	36
76	182ORF076	-3	2316..2426	36
77	182ORF077	2	11858..11965	35
78	182ORF078	-2	7564..7671	35
79	182ORF079	-2	7381..7488	35
80	182ORF080	-2	4372..4473	33

Table 23

Predicted amino acid sequences of ORFs from phage 182

182ORF001

5966 atggcaagaaggatatacaaatgtaaaattgtggctaacgtgccttttgataaacacctatacacacacaagatgggtttaaaact
 1 M A R R Y T N V K L L A N V P F D N T Y T H T R W F K T
 6050 caacagggaacagggaatcgactttaattcgcttctctgcttaacgagaatagagattgcttcttatacaagggtacacacaactc
 29 Q Q E Q E S Y F N S F P V L N E N R D C S Y Q R D T Q L
 6134 gggggaggttttagagtagataaaacacaaagacgccttatatgcttgtaactatctcatcttataaaacgaagaacttatcct
 57 G G V F R V D K H K D A L Y A C N Y L I F K N E E T Y P
 6218 agtaaatggcagtgatgcctttgttactgatattgaatataagaatgacaacacaaagtttcgttacctttgaaattgatgtttta
 85 S K W Q Y A F V T D I E Y K N D N T S F V T F E I D V L
 6302 caaactttatcgctttcgatattgggtatcgagaaagtttcattgcaaaagaacaccctcaactttatttccgaatgggaatacct
 113 Q T Y R F D I G I R E S F I A K E H P Q L Y Y S N G I P
 6386 ttcattaatacaaatgaagagtcgcttgattacggtagagatacacacaacaacaaatgtaacaacttttcatcctaactgatgga
 141 F I N T I E E S L D Y G R E Y T T T N V T T F H P N D G
 6470 gccaattttctgttattctaacaagtgaagcaatgcaggtggagataaggaagataaatcaggaggatcaatagtaggtggc
 169 V N F L V I L T S E A M P V G D K E D K S G G S I V G G
 6554 ccattctctttttctattatttacttctctatcaattcaagtgagggtatatacaacacaaatggggcaggcaatgctaatttt
 197 P S P F S Y L L P I N S S G E V Y K P N G A G N A N F
 6638 ggagatcatggcgctttctacaacgaagaacaccttttaataagatagtcgggatgtatgtaacgctcgatatacaggtata
 225 G E Y M A F L T T K E P F L N K I V G M Y V T S Y T G I
 6722 ccattctgttgatcacgcgaacaaacgggttaaggtataatgcaggaggttcttataagatcatgcttccacactacgtagt
 253 P F I V D H A N K T V R Y N A G G S Y K I M L P T Y A S
 6806 gatccaacagggaacaaatgaaaacattcgctttctttgtgtaaaagaagcaagaacattcgctacctaagaatgatcttga
 281 D P T G T M K T F A P F C V K E A R T F V P K R I D L V
 6890 gggaacggtgataactacttttagagaagcttttccggttaattgtaagggaatcaaaactattttagtaccctattgttataa
 309 G N V Y N Y P R E A F P F N V K E S K L F M Y P Y C L I
 6974 gaaattacagatacaaaaggacatgtaattgactttaagacctaataatcttacaggtggtaaatgagtgatgtgtaaaagg
 337 E I T D T K G H V M T L R P E Y L T G G K L S V Y V K
 7058 tcggttaggaatttctaataaagtgatgatcgagcggatgattatgatgtaagtaactcaaccattattaccaatttaagtac
 365 S L G I S N K V M I E P I D Y D V S N S T I I T N L S D
 7142 aagatggttaactcgataatgatcctaacgatgtaggagtttaaatctgactatgcttctgcattcatgcaaggaaacaaacactc
 393 K M L I D N D P N D V G V K S D Y A S A F M Q G N K N S
 7226 ttgattgctcaagagcaaaacattcgcaatactttcagacatggtatgggaacagtgcaatgagtagcaggaggagcgatcttt
 421 L I A Q E Q N I R N T F R H G M G N S A M S T G G A I F
 7310 tcagccttagcaagtaacaacaccttttgttggttgactaacatcatgggagcaggacaacaaagtaaacacactatggttctgaa
 449 S A L A S N N P F V G L T N I M G A G Q Q V N N Y V S E
 7394 aaagaaaacgggttgaaacctctggcaggtaagtggtgagatatacaaaaatttccagataatgtaaacacagcttggaacaaac
 477 K E N G L N L L A G K V A D I E N I P D N V T Q L G S N
 7478 ttatctttcacaacaggaaactttcaaaactattatcaattgcgcttcaacaacaaatataatagtagtgcacaaagacttgat
 505 L S F T T G N F Q N Y Y Q L R F K Q I K Y E Y A T R L D
 7562 cgttactttcattatgtaggcacaaagcaatcgagtagctacacaaacttcaaaacaaagcaatggaattttcattaaa
 533 R Y S M Y G T K S N R V A T P N L Q T R K A W N F I K
 7646 ttaaaagaacaaatattgtaggcacaaatgagtaacgatgtattaacacgtgtgaaacaaattttagtgaggcggtacgctt
 561 L K E P N I V G T M S N D V L T R V K Q I F S A G V T L
 7730 tggcatcgaatgatgctttgtaattataaccaagcaacggagatgtatag 7780
 589 W H T N D V L N Y N Q D N G D V *

182ORF002

2152 atgattaagaataatactggcgactttgaaacaacactgatctcaacgattgtcggtgatggtcggtggggcgatcgcatata
 1 M I K K Y T G D F E T T T D L N D C R V W S W G V C D I
 2236 gacaaacgttgacaatatgacgtttcggttttagaaatcgattcttttttgagtggtgtaaaatgcaaggcagcacagacattat
 29 D N V D N M T F G L E I D S F F E W C K M Q G S T D I Y
 2320 tcccaacaacgaataatttgacggagagtttatgcttctcaggttattcaaaatggtttcaaatggtgtaagaagcaaaagaa
 57 F H N E K F D G E F M L S W L F K N G F K W C K E A K E
 2404 gatcgaaactctccacactcatatcaaatatgggtcaatgggtatgctttggaaatttgggtggaagttaattacacacaacaa
 85 D R T F S T L I S N M G Q W Y A L E I C W E V N Y T T T
 2488 aaatcaggtaaaaacgaataatctgaacataatattatgatagccttaaaaaatattcctttccagtgaaacaaat
 113 K S G K T K K E K S R T I I Y D S L K K Y P P P V K Q I
 2572 gcagaagcttttaattttcctataaaaaaggcgcaaatagattatcaaaaagaagacatttggttacaacaaacaaagat
 141 A E A F N F P I K K G E I D Y T K E R P I G Y K P T K D
 2656 gaatgggagttatttaagaacgacattcagattatggcgatggcattaaaaattcaattcgatcaaggactaactcgaatgact
 169 E W E Y L K N D I Q I M A M A L K I Q F D Q G L T R M T
 2740 agagggaacgacgcttttaggcgattacaaagattgggttaaaagctacacatggaaaatcaactttcaaaacaaatgggttcttatt
 197 R G S D A L G D Y K D W L K A T H G K S T F K Q W F P I
 2824 tctgtcttttaggtttgataaaagacttacgtataaagcagcaggttctcattgggttaaaacaaagttttcaagggaagaa
 225 L S L G P D K D L R K A Y K G G F T W V N K V F Q G K E
 2908 atagggtgacggcattgtctttgatgtcaactctttgtatccctctcaaatgtacgtaagacctttaccatattggaacacctcta
 253 I G D G I V F D V N S L Y P S Q M Y V R P L P Y G T P L
 2992 ttctcgaagggaataacaaacgaacacgaactatcgctgtacattcaaaatatacaagtaagatttccgttttaaggaggggt
 281 P Y E G E Y K P N N D Y P L Y I Q N I K V R F R L K E G
 3076 tatattccaaccattcaagtaagcaaaagttcattattcattcaaaacgaatattcttgatcaagtgtaaaacaaagtttaggagtt

309 Y I P T I Q V K Q S S L F I Q N E Y L E S S V N K L G V

3160 gacgaattaatcgatcttactcttacaatgttgacctagaattatttttgaacactacgatatttttagagatacattacact
337 D E L I D L T L T N V D L E L F F E H Y D I L E I H Y T
3244 tacggatatatgttcaaagcttcttctgatattgttcaaagctggatcgataaatggatcgaagtaaaagaacaccaccgaagg
365 Y G Y M F K A S C D M F K G W I D K W I E V K N T T E G
3328 gctagaaaagctaacgcgaaggtatgttaaatagcttctgatgaaagttcggaaacaaacccctgacattacaggaaaagtgct
393 A R K A N A K G M L N S L Y G K F G T N P D I T G K V P
3412 tacatggcgaggacggcattgttctgattgacactaggagaagaagaattaaagagatcctgtttatgttccgcttgattgttt
421 Y M G E D G I V R L T L G E E E L R D P V Y V P L A S F
3496 gtgacggctgggttagatatactaccattacaacgcctcaaaaatgtttgatcgcatattttattgtgatacagatagcatt
449 V T A W G R Y T T I T T A Q K C F D R I I Y C D T D S I
3580 catctagtaggaacagaagttccagaagcaatcgacttgggttgatcctaaaaaacttgggtatttgggggcatgaagcaca
477 H L V G T E V P E A I D H L V D P K K L G Y W G H E S T
3664 tttcaacgagcaaaatrcartcggcagaaaacatagtagaagaattgatggcgaattaaatgtaaaagtgtgctgtatgcc
505 P Q R A K F I R Q K T Y V E E I D G E L N V K C A G M P
3748 gatcgaataaaagagattgtaacttttgacaactttgaaagttgggttttcaagctatggaaagttgctacctaagaacacaca
533 D R I K E I V T F D N F E V G F S S Y G K L L P K R T Q
3832 ggtggcgtggtattagtagacacaaatgtttacatcaataa 3873
561 G G V V L V D T M F T I K *

1820R7003

11305 atggaagaacgaattgatattcaaatgaacaagatgaaagaagaaaatcaaaagaattacctattgcaccctgaaacgaacccg
1 M E E R I D I Q M N K M K E E N Q K N Y L L H P E T N P
11389 aaacgaagttgttttggatgaacattgcattgaaatcaggaagtttcaacaattttgttgacacagaagaattgaca
29 K Q V V F D E T L H G N E N Q E S F N N F V D T R K M T
11473 actacaattgatgaagtgcttatgggttatcgctgacgggtgaacagattgtacaccaatattaaataaattacttgaagaa
57 T T I D V S A Y G V I A D G V T D C T P I L N K L L E S E
11557 aaagcggaatgggtatcacttttatttcttctccttctggaacgtgattcatattatcgcttggtaacacattgaaatgaaa
85 K S E M G I T F Y P P C E R D S Y Y R F A N T I E L K
11641 cgtgatgtacctgtagttactttcttaggatcgggagaaacgacattaaagtttgaacaaatgacggcatttaactgaacatc
113 R D V P V V T F L G S G E T T L K F E T M T A P N V N I
11725 gaaagtttcaaatattgatggttttgcattatggttgcacaaaggcgtcgaagtggttaaggaattttctttaatgatactgc
141 E S P N I D G P A L W L P Q G A Q S G K G I F P N A T
11809 aattacaatcgttttgactttgatttgttctgtaactgtacttttaaatgaaggaacgtatggttgttgtgtagagtaga
169 N Y N R P D F D L F V R N C T L N E G T Y V V V A R G R
11893 ggggttaccatttgaataattgtctatttctcaaatcttcaagcaattatcaaaacagcttttcccgatgtaaatgttatgttg
197 G V T P E N C L F S N I S Q A I I K T A F P D V N G M W
11977 caagggacgatatacaactactaggggtacaggttttagaggtttcttcttgtaaaaaaacacgtattcttttgcacagcgtc
225 Q G N D I N T R G T G P R G P F V K N N R I H F C T A I
12061 attatcgacaatgacgatgattatcagaatgtaatttctgtgaaatttctggttaacacaaatcgaaggtggcgttaagttat
253 I I D N D D D Y Q N V I N F C E I S G N T I E G G V S Y
12145 tatcgaggtatgacgataacttgcattgtccaaaaacaacacattttctagcatacggaaatagaacacgtttgttggattt
281 Y R G Y A H N L H V Q N N N H F L A Y G N R N A L F E F
12229 caagtgatggatcaagcttatattgatgtagattgtttattgtcgttaactcacaagtcgaggggaatgaatagtagcagctatttca
309 Q D V D Q A Y I D V D V Y C R N S Q V E G M N S T A I S
12313 cgttttaattgttgtttagcggacattacggaacttaagattacaggttaaatatattcgttgcaggacattgttatcagcttg
337 R L I V V Y G H Y R N L K I T G K L Y R C Q G H V I T L
12397 tatggcgtggcgttaatttctattgtgacttgatggcacaagaagcacttttagcggacgggttacgggttatctaaacgggt
365 Y G G G V N F Y C D L M A Q E A P L T D G Y R F I Q T A
12481 gacaatcgaggttaactatgattgggtttgttctgctggtttgtctcaattcaacaaagtaaatcacacaaatgatacgaatgataaagca
393 D N R V N Y D G F V V R G L S N S T K V N T P M I Y K A
12565 cctcagactgtttctataatcgttagaatcgatcgtgtaacaggtccaatgcaagtaattgtatatcaactag 12639
421 P Q T V F Y N R R I D H V L T G P N A S N V Y N *

1820R7004

4626 atggtcgacaaaatcacagaacaagatggtcttctcggtccacaaatgtagaaccaccagtagcaattaatgactgctatttataat
1 M A D K I T E Q D V L R A T N V E T P V Q L M T A I Y N
4710 agttcatcactcttcttctcaggcgaacgtacattatgcaaatgacataacatcgaagcgggttggtgacgggatcacacgttta
29 S S S L P Q A N V P M P N A D N I E A V G A G I T R L
4794 gacgtatgaaaaacgaattttatttcaacttttagttgacccgtattggttaagtagttatccgatacaaatcttggcgttaaccc
57 D V V K N E F I S T L V D R I G K V V I R Y K S W R N P
4878 ttgaaatgtttaaaaaaggaaacatgccttttaggtcgaaacgattgaagaatttttgttgacattgacaggaaacataagttc
85 L K M P K K G N M P L G R T I E E I P V D I A Q E H K F
4962 aaccctgacgagtcgtttacaggggtattttaaacagggaagttcccgatgtaaaaaacattgttccacgaatttaactcgtgaaggt
113 N P D E S V T G V F K Q E V P D V K T L F H E I N R E G
5046 tactacaaacaaacgatccaagaagcattggttagaanaagcatttacttcatgggataatttcaatagtttctgttgggtgta
141 Y Y K Q T I Q E A W L E K A F T S W D N F N S F A G V
5130 atgaacgctttatcacaggtgacgaagtaagcgaatttgaatacacagaaattattaatagcaaacacgaagaagaagctta
169 M N A L Y T G D E V S E F E Y T K L L I A N Y Q C E K E L
5214 ttcaagagatcgaaattggcgaattactgaatcaaatcaaatgaatttccgtaagatcaaatcaactcctaacaarta
197 F K E I E I G E I T E S N A K E F I R K I K S T S M X L
5298 gaatttatgagttccgcttacaacgctcaaggagttaaaacatctaccrcaaatctgatcaatcgtttattgacggcgac
225 E F M S S A Y N A Q G V K T S T S K S D Q Y V I I D A D

5382 acagacgcaaccattgacggttgacgttttagcagcggcattcaatatgagtaaaactgactttgtaggacacaaaactcgttatt
253 T D A T I D V D V L A A A F N M S K T D F V G H K I V I
5466 gatgagtttctaaaaaaggcgaagaatcgtcaaatattgtggcagttattgtatagtagtgaatggtttatgatctacgac

320

281 D E F P K K E G E E S S N I V A V I V D S E W F M I Y D
5550 aaattgtacaaaacaagaagtcatacaaccctgaagggttatattggaattattggttgaccaccaccaactatattctact
309 K L Y K T T S L Y N P E G L Y W N Y W L H H Q L Y S T
5634 tctcaattcgggaacgctgttgccttttggtaaatcagcaacaaacgtgcacaaaagttgcttttgcagtgcaacaactagt
337 S Q F G N A V A F V K S A T K P V T K V A F A S A T T S
5718 gttgttaaggatcatctaaagatatcgcatcgacatttacaccagtagaagcaacaaacaaaggagaagttgtttcatca
365 V V K G S S K D I A L T F T P V E A T N Q Q G E V V S S
5802 gcaccagcattgttgaaggcaaccgtaaaacacagcaggttaagcgactgcccgaacgtagaaggcttagaagtcggtcaa
393 A P A L V K A T V K Q T A G K A T A V T V E G L E V G Q
5886 tcattagtaacattcacagctatcgagggtcaacaagcaacgggttcttgcgttacttctgactaa 5954
421 S L V T F T A I G G Q Q A T V L V T V T S D *

182ORF005

12651 atggcaactcttacaatgaacaatagctagaggacaaacactcgctaaaatactttcaaatatggctataataaaaattca
1 M A T L T N E Q I A R G Q T I A K I L S K Y G Y N K N S
12735 caagtaggagttgtcgccaatctccattgggaatcggtgttgaacccgaacagcaatgaatattggtgggcnccgcatggg
29 Q V G V V A N L H W E S A G L N P N S N E Y G G G G Y G
12819 ttaggctcaatggagcctaaaagcaatctttatcgccaagcacaaaatttgggggtgtctaatgctaaagctgaacgttggaa
57 L G Q W T P K S N L Y R Q A Q I C G L S N A K A E T L E
12903 ggtcaagcagagatcatcgctcaaggggataaaacaggtcaatggatggataacacactgtttctctcgaggttataactaac
85 G Q A E I I A Q G D K T G Q W M D N T P V S S A G Y T N
12987 cctcagaccctttcagcatttaacaaatctgcaaatattgatgttgcataaattttatgtgctcactgggaacgcccgtgt
113 P Q T L S A F K Q S A N I D V A T I N F M C H W E R P G
13071 aaacttcataatgaagaagcttgatcttgacacagctttagtaagcatattgacggttagcgttggcggtggcgttaaacgt
141 K L H I E E R L D L A Q A Y S K H I D G S G G V K R
13155 tgcattggaaccccaatcaagaatacaaatcttgatcctaaaagtttcatgagtggaacacttttggcagcatgagcaggaac
169 C Y G T P I K N T N L D P K S F M S G Q L F G T H A G N
13239 ggcagaccaataatttccatgatgggttggacttgggttcaattgatcacctggcaatgaatgattgctgttgcgtgga
197 G R P N N F H D G L D F G S I D H P G N E M I A C C D G
13323 acagtaacacatgttggacaatgggagcattagagcgctattttgtgataaatgatgggtactacaatcgcttacaagaa
225 T V T H V G T M G A L R A Y F V I N D G T Y N I V Y Q R
13407 tttagttataaccagtcataataaaaggttaaaagttggcgacaaagtttaagaacggaacagtttgcgcaatcgtgacgagat
253 F S Y N Q S N I K V K V G D K V K N G Q V C A I R D N E
13491 catttacatttaggttttactaaaaagattttatgactgcgttaggatcttctttcatagatgatggaacatgggaagaccct
281 H L H L G F T K K D F M T A L G S S F I D D G T W E D P
13575 ttgaagtttttagggcaatgttttggagatggagatactggcgagataatgacgataacaataaggatcaaaatgatcttatt
309 L K F L G Q C F G D G D T G G D N D D N N K D K N D L I
13659 tatctattgctatccgatgcctgaatgggtggaaatttaa 13700
337 Y L L L S D A L N G W K F *

182ORF006

14995 atgacaaatagcttaggcgttaaaactgaagagaaaaacttatactataaccctaacaatgcttttaggttttaattgcctaattg
1 M T N S L G V K L E E K N L Y Y N P N N A L G F N C L M
15079 ttgtttgtaatagcgccagctgttataggttaaaacttatggttataaaaaatttgttgaatcgctttattaaacacggcgaa
29 L F V I G A R G I G K T Y G Y K K F V V N R F I K H G E
15163 caattttatttaagaagattcaaaacagaacttaaaaagattcctcaatttttcaaaacaatggcgaaaagaatttccgtgat
57 Q F I Y L R R F K T E L K K I P Q F F K T M A K E F P D
15247 cataaaactgaagtaaaaggaaaagaattctattgtgatgataaattaatgggttgggtgttccacttagtacctggggaatt
85 H K L E V K G K E F Y C D D K L M G W A V P I L N E
15331 gaaaaatcctaatgaatcccgaggttcgtacaattttgttgatgagtttttaattgagaatcaaaatcacttattacca
113 E K S N E Y P E V R T I L F D E F L I E K S K I T Y L P
15415 aacgaagctgaagccttattgaacatgatggaacgggttttccgaagcgtacaaatacaagatgtgttatgttgatgaatgca
141 N E A E A L L N M M E T V F R R R T N T R C V M L S N A
15499 actagtgtagtgaaaccttatttctgtatttcaatcgcagcagatttgaataagcgttttaactatatacaagatcgaggt
169 T S V V N P Y P L Y F N L Q P D L N K R F N L Y Q D R G
15583 atattgatgaattgtgtgattcaaaagactttgcagaagtgagagagaaacaccttttggtagattgattcgtggacagaa
197 I L I E L C D S K D F A E V K R E T P F G R L I R G T E
15667 tacgaagatttttagtatacaaatgagttgtcnaatgatagtgatacgtttattgaaaagagaagtaaaatagtagtttctta
225 Y E D F S I N N E F V N D S D T P I E K R S K N S S F L
15751 tgcgccattgcttttgaagggaatccttgggtattggatagacgctgaaacaggttgggtctatgtgagttatgattatcaa
253 C A I A F E G K I F G Y W I D A E T G C V Y V S Y D Y Q
15835 ccaatacaaatcattttttagcaatgactacgaagaccatgaagaaaatagattgctgatgaaaaattggcgaaataattat
281 P N T N H F Y A M T T K H E E N R L L M K N W R N N Y
15919 tatctttcaacagtgccgaagcattcaagaatagttatctgcgggttgataacattgttattagaatttacattatgattg
309 Y L S T V A K A F K N S Y L R F D N I V I K N L H Y D L
16003 tttaataagatgaaaatctggttaa 16026
337 F N K M K I W *

182ORF007

7795 atgagtagacgaaaagggtgcaggacttgctagaaaataaccgttatcacgcaaaaagcagaccttatccaaatgaacccctattca
1 M S R R R K G A G L A R N N R Y T A K S R P Y P N E P Y S
7879 agtgatgtagaagaatcagctactatgaacattatcgtagacaactcagctccttacggtttcagtttggatggggaat
29 S D V E E I S Y Y E H Y R R Q L T L L T F Q L T F E W E N
7963 ttgccaaaatcaattgacctcgttatttagaaattgctttacacactaatgggttatcttgggttctttaaagaccctacactt
57 L P K S I D P R Y L E I A L H T N G Y L G F P K D P T L
8047 ggggtcatggtttgcgcaggggcagaagatgggtcaaatcgatcattatcacacacctattttctttacagcaaacgaagcaatg

321

85 G F M V C A G A E D G Q I D H Y H N P I F F T A N E A M
8131 tatcacaagagatatcctgttttaagatatgatgatgatgataaatcaaatgtatcatgtgtataataatgacttgaaa
113 Y H K R Y P V L R Y D D D D D K S K C I M L Y N N D L K
8215 gttcctacgttaccagtttcatcgttttgccttagatagtgccgacataaaccagatatcagcagtgatcgagagcgcaa
141 V P T L P S L H R F A L D M A D I N Q I S R V N R R A Q
8299 aaaacacctgtaattattcaactgatgaaaagaaatcttctcattgctcaagcttataaaccasattgacgaaaaataatcag
169 K T P V I I Q T D E K K Y F S L L Q A Y N Q I D E N N Q
8383 gctgtttttgtggataaagatatggagtttgacgaattcttttaattgtatggcaaacaaatgctccatgttagtagataaaacta
197 A V F V D K D M E F D E S F N V W Q T N A P Y V D K L
8467 cgatcagaattgaacgaagtatggaatgaagtgttaacttttctaggatcaacaatgctacgttagataagactgcacgtgta
225 R S E L N E V W N E V L T F L G I N N A N V D K T A R V
8551 caaacatcagaagttcttatcaacaatgaacagattgaaagttcaggtaacatcttctgtaaaatcaagaaaagagtttgcgat
253 Q T S E V L S N N E Q I E S S G N I L L K S R K E F C D
8635 cgtgtaaatcgtgtcttggcgatgaacttgacggaagattgacgtgaagtttagaacagacgctgcagaaattacaaactg
281 R V N R V F G D E L D G K I D V K F R T D A V R Q L Q L
8719 gcggcaggtcaatcaaaaaagaccagatgagtgagggttgccaagtgcctactaa 8775
309 A A G Q S K K D Q M S G G L P S A T *

182ORF008
14105 atgatgaatggattgatattctctagttatcaaacaggattgatctttcaaaagttccatgcgattttgtaaatattaaagca
1 M N G I D I S S Y Q T G I D L S K V P C D P V N I K A
14189 acaggcggaacaggttatgtaaaccttgattgtgacggagcatttcaacaagctttgctctttaggtaaaaagattgggtgtgat
29 T G G T G Y V N P D C D R A F Q Q A L S L G K K I G V Y
14273 cattttgcgcatgagaggggttttagaaggtacacctcaacaagaagcgcaattcttttagataaatatagggttacattggt
57 H F A H E R G L E G T P Q Q E A Q F F L D N I K G Y I G
14357 aaagctgttcttattcttgactttgaagggcctcaaatcagaaaagattgaaattgggagaaagcatttcttgattatgtttataat
85 K A V L I L D F E G S N Q K D V N W A K A F L D Y V Y N
14441 aaaacagggcttaagcatggttttatcgtatcacgaaacctcaatacaactgatttttctagattgcaaaagcgattat
113 K T G V K A W F Y T Y T A N L N T T D F S S I A K G D Y
14525 ggtttatgggttgcgaatatggaatcaaatcaaccacaggctactctcaaccagcgccacctaaacaaataatatttcccaatt
141 G L W V A E Y G S N Q P Q G Y S Q P A P P K T N N P P I
14609 gttgcctgttttcagtttacaagtaagggagctttaccaggatacaacggcaactcttgatttgaattgtttctatggcgatgggt
169 V C C F Q F T S K G R L P G Y N G N L D L N V F Y G D G
14693 aatcatcgggatctgtatgtagtataaaacaggatcaaatgttctcctgaaaaataaatatttgacgcccacagtgatgag
197 N T W D L Y V G K K Q D Q I V P P E N K I F D A T S D E
14777 tttattttcactcttacaacaggttagcacagcgtgttttattttgacgggagaacgatctttgaattgtctgatccaacacaa
225 F I P T L T T G S T S V F Y F D G E T I F E L S D P T Q
14861 ctgcgcatattagaggaacatacaatcatgttctggaagaaatcccatcaatggtgagacacctgacacaaatttgatatt
253 L D H I R G T Y N H V H G K E I P S M V W T P E Q P D I
14945 tacttaaaatgtatgaanagaaccagtatataaacag 14983
281 Y L K M Y E K K P V Y K *

182ORF009
8765 gtgctacttaaacgttatattgaaagtttcacttattaccaacctgaattatctcgaaaagaacgtattgaaagttggcgcaaaa
1 L L K R Y I E S P T Y Y Q P E L S R K E R I E V G R K
8849 caattgtttgattttgattatccgttttatgacgaacaaaacgagcagaatttgaacaaaatttatcaatcacttttacttg
29 Q L F D F D Y P F Y D E T K R A E F E T K F I N H F Y L
8933 agagatagggctcagaaaacgatgggtcattttaagtttaattcttgacgaatatttaaatctaaacatgcccatttggaataaa
57 R E I G S E T M G S F K F N L D E Y L N L N M P Y W N K
9017 atgttccatcaaatcttgaagagtttccgatttttgatgacatggactacaccattgatgagaaacagaaaattgttaaatgag
85 M F L S N L E E F P I F D D M D Y T I D E K Q K L N E
9101 attgatacaaacatcaaacggaatcgtgatgaatcgaagaacaaacgaagcaagtagatcaaacagacacagaaacaaaat
113 I D T N I K A N R D E S K N Q T K Q V D Q T D N R N K N
9185 acagtgacacaggaacacaggtattctttctcaaggaaacacttatcacagacacccctcaaaaagatttgagaaattggcagcaat
141 T R D T G T T D S F S R N T Y T D T P Q K D L R I A S N
9269 ggagatggaaacaggtgtaattcaattatgcaacaaatcacagaagatttgagtaagaaacaaacagctccacagggcgttgaa
169 G D G T G V I N Y A T N I T E D L S K E T T S T G V E
9353 acaaaacacgacaaaacaaatcaaaatcacgaagcaatgcttctgaaaaagaaacaaagacacagacattataaagatcaa
197 T N N D K T N Q N T R S N A S E K E T K N T D I N K D Q
9437 aatcaaaacaaagatcagattacacgatataaagggtaaaaagggaaacactgattatgtgacttactcgaaaaatattcgtaga
225 N Q T K D T I T R Y K G K K G N T D Y A D L L E K Y R R
9521 agtgttttgagaattgagaaaatgatcttttagagaaatgaacaagggaaggcttatttctcctgtttatggaggaggtag
9601
253 S V L R I E K M I F R E M N K E G L F L L V Y G G R *

182ORF010
1310 ttgacgctaaagaatatcaagaatgatagagccaagtttagagaaaatctacggttaaatctaacaagctcgtaaaaaatacaat
1 L T V R I S K N D R A K L E K I Y G K S N K A R K K Y N
1394 cgtttaagacaaaaaggagttgaggaaaaggcaacttccaactgttccaacatcaagaaaagacttattgactacgttaaatca
29 R L R Q K G V E E R Q L P T V P T S K K R L I D Y V Y
1478 acaaatatgagtcgtagtgatttttaacaagattgttagacgagttggtagattttgcacaaccttacaacgagaatttaoatttt
57 T N M S R S D F N K M L D E L V D F A Q P Y N E N Y I F
1562 gagatcaacaagcgaatgttgcaatctcaagagcgcaaatcaagaagcgcaaatcaaaacagagcaagctcaaaaagcgaaa
85 E I N K R N V A I S R A G I K E A Q I K T E Q A Q K A K
1646 gaagaacactacaagagcttaacaagttgaagttgaagagccacagaaacacaaattgtcacaccaacttttaacagag
113 E H Y K E L N K V E V K K P T E N T I V T P T I L T E
1730 ttagggtgctgacttaccttttcaagcaataccagatttcaatattgacgctttcacttctccagaaggagttcagttctattta

141 L G A D L P F Q A I P D F N I D A F T S P E G V Q S Y L
1814 gaaaatataggaacaaagacgaacaaatattttgacgaagagacaaactttartacgacaatttcagacaagcgatgtttact
169 E N I G K Q D E Q Y F D E R D Q L Y Y D N F R Q A M F T
1898 attttcaattcagacgctgacgatattgttcgtttacttgactcaatggggttgatctattttatgaaaacatatgttagtaac
197 I F N S D A D D I V R L L D S M G L D L F M K T Y V S N
1982 ttcttagacatgaaccttgactacatttatgacgaagcagaagtacacagaaaaaagaacaaagttttagtaagattgcaaaa
225 F L D M N L D Y I Y D E A E V Q Q K K E Q V Y S K I A K
2066 gtgatcgagctgaaacaggtggagaagtcctcatataacccacgaagaacatcacaaattttagtaaacaggagaagaa
253 V I E S E T G G E V P S Y N P T K N I T I N S E T G E E
2150 ttatga 2155
281 L *
182ORF011
9607 atggtagatttttaaccccgacaagcggtttgacgggtttaccgctgtattcaagaacgcttttagcaaatatcctcatactgaa
1 M V D F N P D K R F D G L P A V F K E R F S K Y P H T E
9691 tacagatatgaattactattagatgaagaagtatcggttttaattgccttatctgaatgaagttggtgctttagttaatgatag
29 Y R Y E L L L D E E V S A L I A Y L N E V G A L V N D M
9775 agtgggtatttttaattactttatcgaacattttgttgagaagttagaagagatcacaaatgacacactcaaaaaatgggtgctt
57 S G Y L N Y F I E H F V E K L E E I T N D T L K K W L S
9859 gacggtacgttagaaaaatttaataatgatactgttttgcgaatttatcaaaagaaatcaaaagattacaaatcttggttgct
85 D G T L E N L I N D T V F A N Y I K E I K R L Q I L V A
9943 gaaacacgtgctaacagtgatattcttttgacaaaaataaacggatgttgctgatgacgaacattttggtataagatt
113 E T R A N S V N I L L T K N K P D V A D D R T F W Y K I
10027 caacgcgaataactgattatggagccgacccattgacacgttaactgttgcgaatcaataaagtttagtggtggaataacc
141 Q R D N T D Y G A D P I D T L R I V A I N K V S G W N T
10111 gctacaggagatattttatcttaacattaaaggaacggaggtgtataa 10158
169 A T G D I Y L N I K G T E G V *
182ORF012
10872 atggcaaatataaattattcaaatgaaggatagcaaatatatttatccaagtgttcgagcagaaaaacttgtagatttg
1 M A N K N I Q M K D S N D N N L Y P S V R A E N L L D L
10956 accagtcgtgctgaattacaattgacaaattgtcaatttatcgcagctggtgataaaaacaaatgcaatctcttattcgcgtgca
29 T S R A E L T M T N C Q L Y A A G D K T N A I S Y L G A
11040 gtgagttatgctcgaaggtatgataaagtttactgaaagtttgacaaacccctgtgatcacacgctaccacaggttttagacca
57 V G M L E G M I K F T E S L T N P V I T T L P E G F R P
11124 ataagaacaaaactgattgtgttttcgcaaaattttacacacaaatccaacagatacaaaagaaatgggtttagtataatc
85 I R T K R I G C F A K Y Y T P N P T D T K E M V Y V S I
11208 acaccttgccgaagtaactgataatgacaaatgaggaataatcgaatattcttatccctagataaattcgcttttccctctaaaa
113 T P D G K V T V N D N V G K I E Y L S L D N C V F P L K
11292 taa 11294
141 *
182ORF013
10456 atggcagataaaaaattattcaaatgcaggataaagatcataatcggttaattgctgttacaattgctaaaaatgttctaacaggc
1 M A D K N I Q M Q D K D H N R L M P V T I A K N V L T G
10540 gactctaatcttgaattagtttaattgctgaataaagagttacgctagtgaaagctaaaaacacttgacacaaagctaaagaaact
29 D S N L E L V N A E I R G N A S E A K T L A Q K A G I I
10624 gctgctggtttgtcaacagaaattgacacagtaacatcaaccgcaaatcaagcgttgacgaaggtggtacagcacacaacacc
57 A A G L S T E I D T V T S T A N Q A L T K A G T A Q Q T
10708 gcagaaacagcgaacacacagcaaacagatcagcgaggttgcaacggcagctaaaaacacagctgattcagcacaaaaaggt
85 A E Q A K T T A N S I S A V A T A A K N T A D S A Q K S
10792 gcaactgatctagctgttgagtagtaagcagtttagagggacacagcaatacatatactgtattaccatag 10860
113 A T D L A V R V S S L E D T A I Q Y T V L P *
182ORF014
13716 atgatagaatatcacacaatggttggcagatgataatcatcttgggtttgatttatggttaattggttgcaatgatt
1 M I E Y I T Q W L A D D N H L V Y G L I I W L M V A M I
13800 atcgattttgtgttagttttacaaattgccaatttaacaagggaatcgacttttagtatttttaagctaaagcaggtatcatt
29 I D P V L G F T I A K P N K E I D F S S F K A K A G I I
13884 gtttaaggtggcagaatggttttagtgggttactttcttctgtagcagtaaaatcggtgacagtaggtattacaatgtatata
57 V K V A E M V L V V Y F I P V A V K F G A V G I T M Y I
13968 acaatggttgggtttgattttatcagaaatttagtatactaggacatatttcagatatcgatgatataataattggact
85 T M L V G L I L S E I Y S I L G H I S D I D D D N N W T
14052 gattatgttaagaagtttttagacggaacactcaacgaagaaggacgatataaatga 14108
113 D Y V K K F L D G T L N R K D D I K *
182ORF015
854 atggaaatcgtaaaaagcacatttgacacacaaacaccagaaggaatgttacaagtattcaatgccacaaacgggggttcaatt
1 M E I V K S T F D T Q T P E G M L Q V F N A T N G A S I
938 ccgttacgtaacgcaattggcgaagtactagaattgaaagatattctagtttactcagacgaagtttctggttttggtggagcc
29 P L R N A I G E V L E L K D I L V Y S D E V S G F G G A
1022 gaacatcacaaagcagaactagtcggtttcttcacagaagatggtaaaacttatcggggtgtatcagcagtagcaacaaatca
57 E P S Q A E L V A P F T E D G K T Y A G V S A V A T K S
1106 gctaaaaacctaattgatagtagtgcacacccatcaacccaaaatttctttgtcgaagggaatcaaacggtgga
85 A K N L I D M M T A N P D I K P K I S P V E G -K S N -G
1190 caaaaatttgtaaatctacaagtggtttcactgtag 1225
113 Q K F V N L Q V V S L *
182ORF016
17033 atgattacaatttatcattatttttaggggttttaactcaactaactaaagatgacaacgatagtttagcgtctatcaagtca
1 M I N N L S L I L E G L N Q L T K D D N D S L A S I K S
16949 gaaataacacaaggaggaacaaatattttatatactgattacgtttacaaaagagttcgtgttaacacatgataataaac

323

29 E I T Q G G K Q L I L Y I D Y V T K E F V L T H D K Y N
16865 tatgtttatcttgatagccattgcattaatcgcgaataacgaaatcaatgaaaagcgttgaaacactatgcggaacaaattgaaa
57 Y V Y L D S H C I N I A I T K S M K S V E H Y A E Q L K
16781 catgacggatataaaacaaattacggacaaatag 16749
85 H D G Y K Q I T D K *
182ORF017
154 atgaaatattcactacaacaataatagatgaaatcaacaattttcagaattagattaaaaagcagtaactagaggaattg
1 M K Y S L Q Q I D E I K S T I F R I R L K R H E L E L
238 gtggacgaagtaaacgataattgctaaagatccggaggaaagatatcttttatcgtttattacacagaagaagaacgtttgttt
29 V D E V N D I A K D P E E R Y L L S F Y Y T E E E R L F
322 gaaattccctctgcaagattaatagattattacaacgaaaagatcacaaatctgaaatcggaatcatatcactcgaaaaaaga
57 E I P S A R L I D Y Y N E K I T N L K S E I I S L E K R
406 ttacaaaaactagtaaaataa 426
85 L Q K L V K *
182ORF018
16737 atgattgcacgaacattcaagaacaccgcgaactaattgaatgggtacgtttctactgtaaacgttaacctttcagacaatgaa
1 M I A R T P K E H R E L I E W L R F Y C K R N L S D N E
16653 aaaatagagatcatagagggactttacaagatttcgacgttccggaaaataaatatcacggaacttttgtaactcattcaacg
29 K I E I I E G T L Q D F D V P E I N I T E L L L T H S T
16569 ctattaccgaatcgagtcatttaacattcttgaaaagattgtcaggaatgaaattagtaacttcacgttaaaagttggt
57 L L P E S S Q F N I L E K Y C Q A M K L V T S Y V K V G
16485 tctcgctatcagttagcgttacaaataccaaaaggctatttaaggaggtggaataa 16429
85 S R Y Q L A L Q I P K G Y L K E V E *
182ORF019
4323 atggaataaagaacatgaatcaatttttaattggtattcttgaaaagtgctcacagacggtgaagcaagatcasagattgtagaa
1 M E I K E H E S I L N G I L E S V T D G E A R S K I V E
4407 catcttgaaagcattgcgagaagactacggagcaaacactgaagctttgacatcagcaaatagcacacttggaagtttaagaaa
29 H L E A L R E D Y G A T T E A L T S A N S T L E K L K K
4491 gataacgaagcgttggtattttcaaaactcaaaattgttccgagaacgagcgatcgtagaaccagcagaaaaataacgaaccagaa
57 D N E A L V I S N S K L F R E R A I V E P A E N N E P E
4575 acagaccagaatattacactagacgatttaggaatttaa 4613
85 T D Q N I T L D D L G I *
182ORF020
10158 atggcagacattagaacacaaactaaagtgaaatggatcagacaattttttccaatttcasaagcgttaattatgact
1 M A D I R T Q L T S E D G S D N L F P I S K A V N I M T
10242 aatagcgggtacgaatgtagaagggaattgggtacactcaaaacaaatgacgaacaaatgaatacctcagttcaaaaatgctgta
29 N S G T N V E G E L G T L K Q N D E T M N T S V Q N A V
10326 gttactgccaatcaagcaaaagattctgtagctgaattaaatgtaaatgttggttaactaaccaatcgaataacaacattagag
57 V T A N Q A K D S V A E L N V N V G K L T N R I T T L E
10410 agtacagtggttaattcttgatggtattcgttatgtagaggtgtaa 10454
85 S T V A N L D G I R Y V E V *
182ORF021
17339 atgaacaataaatcattaatagctgaaaaaggagaggtatctctacttcaacccctttaaagagtggtgatatgaattatcatatc
1 M N N K S L I A E K G E V S L L H P F N E W D M N Y H I
17255 atagataccgaaaacaaataaacattatcttattgatattgatgaggtaggcgatgaggaatattgtttgttatcttttgaaagaa
29 I D T E N N K H Y L I D I D E V G D E E Y C L L S F E E
17171 ctaaaggaattagatatggatctttttccgagttattcatggaataactacagaataacatattaa 17106
57 L K E L D M D L I S E Y S W K T T E I T Y *
182ORF022
12868 gtgggtgtgtctaatgctaaagctgaaacgttggaaggtcaagcagagatcatcgctcaaggggataaaacaggtcaatggatgg
1 V G C L M L K L K R W K V K Q R S S L K G I K Q V N G W
12952 ataatacacctgtttctctgtcaggttatactaacctcagaccctttcagcatttaaaccaatctgcaaatgaaattgattggtcta
29 I I H L F L L Q V I L T L R P F Q H L N N L Q I L M L L
13036 caattaattttatgtgtcactgggaacgcccctggttaacttcataatcgaagaagacttgatcttcgacaaagcttatagtaagc
57 Q L I L C V T G N A L V N F I S K K D L I L H K L I V S
13120 atattgacggttagcgtggtggtggtggttaa 13149
85 I L T V A V A V A *
182ORF023
12189 atggtgtgtgtgttttgacatgcaagttatgcgcatatcctcgataataacttacgccaccttcgattgtgttaccagaaatttc
1 M V V V L D M Q V M R I S S I I T Y A T F D C V T R N F
12105 acagaataataattacattctgataatcatcgctattgtcgataatgatcgctgtaaaaaatgaatacgtgtgtttttcacaaa
29 T E I N Y I L I I I V I V D N D R C T K M N T V V F H K
12021 gaaacctctaaaacccctgtacccttagtattgatatcggttcccttgccacataccatttacatcgggaaaagctgttttgataat
57 E T S K T C T P S I D I V P L P H T I Y I G K S C F D N
11937 tgcttgagagatattagagaatag 11914
85 C L R D I R E *
182ORF024
6174 atgcttgtaactatctcatcttttaaaacgaagaacttatcctagtaaatggcagtagcctttgttactgagattgaataa
1 M L V T I S S L K T K K L I L V N G S M P L L L I L -N I
6258 agaattgacacacaaagtttcgttacctttgaaattgatgttttacaaacttatcgtttcgatattggtatcagagaagtttca
29 R M T T Q V S L L P L K L M F Y K L I V S I L V Y E K V S
6342 ttgcaasagaacacccctcaactttattttcgaatggaatacctttcattaatacaattgaagagtcgcttgattacggttagag
57 L Q K N T L N F I I R M E Y L S L I Q L K S R L I T V E
6426 aatacacacacaaatgtaa 6446
85 N T Q Q Q M *

182ORF026

1820RF027

1820RY028

1820RF029

182ORF030

1820RF032

1620RF033

1820RF034

5160 gtgtttatctactctaaaaaactccccgagttgtgtatccctttgataagaacaatctctattctcgtaagaacaggaaacga
1 V F I Y S K N S P E L C I P L I R T I S I L V K N R K R
6076 attaaagtagcagttctgttccctgttgagttttaaacpatctctgtgtgtatagggtgttatcaaaaggcagcttagccaacaa
29 I K V R F L F L L S F K P S C V C I G V I K R H V S Q Q
5992 ttctacattgtatacctctctgccataattgtcctccttag 5951

325

57 F Y I C I P S C H N C P P *
1820R035
15758 atggcgcatagaactactacttttttacttctctttttcaataaacgtatcactatcattgacaaactcattgtgatactaaaa
1 M A H K K L L F L L F S I N V S L S L T N S L L I L K
15674 tcttcgtattctgttccacgaatcaatctacaaaagggtgttctctcttctcacttctgcaaagtcttttgaatcacacaattca
29 S S Y S V P R I N L P K G V S L F T S A K S F E S H N S
15590 atcaatatacctcgatcttga 15570
57 I N I P R S *
1820R036
2315 atgtctgtgctgccttgcattttacaccactcaaaaaagaatcgatttctaaaccgaacgtcatattgtcaacgttgcctata
1 M S V L P C I L H H S K K E S I S K P N V I L S T L S I
2231 tcgcatacgcctccacgaccatacagacaatcggtgagatcagttgtgtttcagaagtcgacgtatatttcttaatacataatt
29 S H T P H D H T R Q S L R S V V V S K S P V Y F L I I I
2147 cttctcctgtttctgaattaa 2127
57 L L L F L N *
1820R037
12280 gtgagttacgacaataaacatctacatcaatataagcttgatccacatcttgaaactcaacaaagcgtttctatttccgtatg
1 V S Y D N K H L H Q Y K L D P H L E T Q T K R F Y F R M
12196 ctgaaaaatgggtgtgttttgacatgcaagttatgcgcataatcctcgataataacttacgccaccttcgattgtgttaccag
29 L E N G C C F G H A S Y A H I L D N N L R H L R L C Y Q
12112 aaatttcacagaaattaa 12095
57 K F H R N *
1820R038
14769 gtgatgagttatttttactcttacaacaggttagcacaagcgtgttttattttgacggagaaacgactctttgaattgtctgac
1 V M S L P S L L Q Q V A Q A C F I L T E K R S L N C L I
14853 caacacaaactcgatcatatttagaggaacatacaatcatgttcgataaagaatcccatcaatgggtgacacactgaacaat
29 Q H N S I I L E E H T I M F M E K K S H Q W C G H L N N
14937 ttgatatttacttaa 14951
57 L I F T *
1820R039
9992 atgttgctgatgacacatttttggtataagattcaacgcgacaataactgattatggagccgatcctattgacacgttacgta
1 M L L M I E H P G I R F N A T I L I M E P I L L T R Y V
10076 ttgttgcaatcaataaagtttagtggtggaataccgctacaggagatatttatttcaacattaaaggaacgggggtgtataat
29 L L Q S I K L V A G I P L Q E I F I L T L K E R R V Y N
10160 ggcagacattag 10171
57 G R H *
1820R040
16202 atgagaaaaagatttcgtctacattaacacacccgatccaaaagcaacaaaaaggcgttagcaaaaatcactaacgccaaagaa
1 M R K D P V Y I N T P D P K A N K K A L A K I T N A K E
16118 ccaaaaacaaactatcgagactacaactactatgttcttactattcatctgtaataagaactaatcgtggtgacttacta
29 P K Q N Y R R L Q L L C Y L L F I I V I E L I V V A L L
16034 aaatag 16029
57 K *
1820R041
3886 atggaactatataaagcaatgtttatcgtagtgatgaaggtactattgacggttacgatactgaacactatgtagatatttct
1 M E L Y K A M F I V R D E G T I D G Y D T E H Y V D I S
3970 ttacatgactttgaagaatataatgaaaaagaacacgtgaaattgaagcagtaaacattagtaaaaaacaggaatttaaaaaa
29 L H D F E E I Y G K E T R E I E A V T L V K T G N L K K
4054 taa 4056
57 *
1820R042
10832 gtgtcctctaaactgcttactcgaacagctagatcagttgcacttttttgctgtaacagctgtgttttagctgccgttgca
1 V S S K L L T R T A R S V A L F C A E S A V P L A A V A
10748 actgcgtgatactgtttgtgtgttttcgctgtgttctcggtttgtgtgtgtaccagccttcgtcaacgcttga 10671
29 T A L I L F A V V F A C S A V C C A V P A F V N A *
1820R043
10652 gtgtcaatttctgttgacaaaccagcagcagtttctttagctgtgttgcaaggttttagcttactagcgttacctcttatt
1 V S I S V D K P A A V S L A C C A S V L A S L A L P L I
10568 tcagcattaactaattcaagattagagtcgctgttagaacatttttagcaattgtaacaggcattaaacgattatga 10491
29 S A L T N S R L E S P V R T F L A I V T G I K R L *
1820R044
6457 atgaaaaagttgttactttgtgtgtgtattctctaccgtaatacaagcgactcttcaattgtattaatgaaaggtattccatt
1 M K S C Y I C C C V F S T V I K R L F N C I N E R Y S I
6373 cgaataataaagttgaggtgttcttttgcaatgaaactttctcgatataccaatatcgaaacgataagttgttaa 6299
29 R I I K L R V F F C N E T F S Y T N I E T I S L *
1820R045
6729 atgaatggtatacctgtatagcaggttacatcacactcccgactatcttatttaaaaaagggttcttctgttgtaagaacgccatg
1 M N G I P V Y D V T Y I P T I L F K K G S P V V R N M
6645 tactctccaaaattagcattgcttgcctcccttgggtgtatcctccctcactgaattgataggaagtaataa 6571
29 Y S P K L A L P A P F G L Y T S P L E L I G S K *
1820R046
2372 atggtttcaaatggtgtaagaagcaaaagagatcaaacattctccacactcatatcaaatatgggtcaatggtatgctttgg
1 M V S N G V K K Q K K I E H S P H S Y Q I W V N G M L W
2456 aaatttgggaagtttaattacacaacaacaaatcaggttaaaacgaaaaagagaattctcgaacaataa 2527
29 K F V G K L I T Q Q Q N Q V K R K K R N L E Q *

182ORF047
13353 atgtctccattgttccaacatgtgttactgttccatcgcaacatgcaatcatttcattgccagggtgatcaattgaaccaaagt
1 M L P L F Q H V L L F H R N M Q S F H C Q G D Q L N Q S
13269 ccaaacatcatggaaattatttggctgctgcttctctgcatgctgcaaaaagtgttccactcatga 13201
29 P N H H G N Y L V C R F L H A C Q K V V H S *
182ORF048
3395 atgtcagggtttgttccgaactttccatacaagctatttaacatacctttggcgtagcttttctagcccccttcgggtgtgttc
1 M S G F V P N F P Y K L F N I P L A L A F L A P S V V P
3311 tttacttcgatccatttatcgatccagcctttgaacatatcacaagaagctttgaacatatatccgtaa 3243
29 F T S I H L S I Q P L N I S Q E A L N I Y P *
182ORF049
1578 atgttgcaatctcaagagcgcaaatcaagaagcgcaaatcaaacagagcaagctcaaaaagcgaaagaacactacaaag
1 M L Q S Q E R K S K K R K L K Q S K L K K R K K N T T K
1662 agcttaacaaagtgaagttgaagaagccacagaaaaacaaattgtcacaccaactatttttaa 1724
29 S L T K L K L R S P Q K T Q L S H Q L F *
182ORF050
8012 atgggtattcttggtttctttaaagaccctacacttgggttctatgggttgcgcaggggcagaagatgggtcaaatcgatcattatc
1 M V I L V S L K T L H L G S W F A Q G Q K M V K S I I I
8096 acaaccctatttctttacagcaaacgaagcaatgtatcacaagagatctctgttttaa 8155
29 T T L F S L Q Q T K Q C I T R D I L F *
182ORF051
9390 atgcttctgaaaaagaacaaagaacacagacattataaagatcaaaatcaaaccaagatacgattacagatataaaggta
1 M L L K K K Q R T Q T L I K I K I K P K I R L H D I K V
9474 aaaagggaaacactgattatgctgacttactcgaaaaatctcgtagaaggttttga 9530
29 K R E T L I M L T Y S K N I V E V F *
182ORF052
4096 gtgatagttgacaagagtc aaatttggcgagattggcgcaatgtacacgtgaaatctcgtgcgctcccgtaaagttatggacac
1 V I V D K S Q I W R D W A N V H V K Y R A L P L S Y G H
4180 ataaacgttttgaccgtcaaccaatcgcaaaaaccttttaggagtagcccttaa 4233
29 I N V L T V N Q S Q K P F R S S P *
182ORF053
15656 gtggaaacagaatacgaagatttttagtatcaacaatgagtttgcattgatagtgatacgtttattgaaaagagaagtaaaaaata
1 V E Q N T K I L V S T M S L S M I V I R L L K R E V K I
15740 gtatgtttcttatgcgccattgtctttgaagggaatactttgggtattgtag 15793
29 V V S Y A P L L L K G K S L G I G *
182ORF054
8136 gtgatacattgcttctgttctgtaagaaataggggttgataatgatcgatttgaccatcttctgccctcgcaaacat
1 V I H C F V C C K E N R V V I M I D L T I F C P C A N H
8052 gaacccaagtgtagggtctttaaagaaccaaagataaacatttagtgtaa 8002
29 E P K C R V P K E T K I T I S V *
182ORF055
8324 atgaaaagaatacttctcattgtctacaagcttataaccaaatgacgaaaataatcaggctgttttgggataaagataggg
1 M K R N T S H C Y K L I T K L T K I I R L F L W I K I W
8408 agtttgacgaatcttttaattgtatggcaacaaatgtccatagtag 8455
29 S L T N L L M Y G K Q M L H M *
182ORF056
6549 gtggcccatctcttttcttatttacttctcatcaattcaagtgggaggtatcaaaacaaatggggcaggcaatgcta
1 V A H L L F P I I Y F L S I Q V G R Y T N Q M G Q A M L
6633 attttggagtagatagggctttcttacaacgaagaaccttttttaa 6680
29 I L E S T W R P L Q R K N L F *
182ORF057
8264 atgtccgcatatctaaagcaaaacgatgtaaaacttggttaacgtaggaactttcaagtcattattatcaaacatgatcatttt
1 M S A I S K A K R C K L G N V G T F K S L L Y N M I H P
8180 gatttatcatcatcatcatcatatctttaaacaggatattcttctgtga 8133
29 D L S S S S S Y L K T G Y L L *
182ORF058
5176 gtgtattcaaatcgcttacttctgctacctgtgtataaagcggttcattacaccagcaaacgaaactattgaaattatcccatgaa
1 V Y S N S L T S S P V Y K A F I T P A T K L L K L S H E
5092 gtaaatgctttttctaaacctgcttcttgatcggttggtag 5048
29 V N A F S N H A S W I V C L *
182ORF059
15876 atggctcttctgtagtcattgcataaaaaatgatttgtatttgggtgataatcataactcacatagacacaacctgtttcagcgtc
1 M V F R S H C I K M I C I W L I I I T H I D T T C F S V
15792 tatccaataaccaagattttcccttcaaaagcaatggcgcataa 15748
29 Y P I P K D F P F K S N G A *
182ORF060
15404 gtgatttttgaatttctcaattaaaaactcatcaaaaattgtacgaacttcgggatattcatttagatttttcaattccccac
1 V I P D P S I K N S S N K I V R T S G Y S L D F S I P H
15320 gtactaagtggaacagcccaaccattatattatcatcacaatag 15276
29 V L S G T A Q P I N L S S Q *
182ORF061
2102 atgaggggacttctccacctgttcagactcgatcacttttgcaatcttactgtaaacttgttctttttctgttacttctg

327

1 M R G L L H L F Q T R S L L Q S Y C K L V L F S V V L L
2018 cttcgtcataaatgtagtcaagggttcctgcttaagaagtactaa 1974
29 L R H K C S Q G S C L R S Y *
1820RF062
1992 atgtcttaagaagtactaacaatattgttttcataaatagatcaagccccattgagtcagtaaacgaacaatatcgtagcgtct
1 M S K K L L T Y V F I N R S S P I E S S K R T I S S A S
1908 gaattgaaaatagtaaacatcgcttctgtctgaaattgtcgtaa 1867
29 E L K I V N I A C L K L S *
1820RF063
14306 gtgtaccttcttaaacccctctcatcgcgcaaaaatgatacacaccaatcttttacctaagaacaaagcttctgaaatgctcggt
1 V Y L L N P S H A Q N D T H Q S F Y L K T K L V E M L G
14222 cacaatcagggtttacataaacctgttccgctgttctttaa 14181
29 H N Q G L H N L F R L L L *
1820RF064
7356 atgatgttagtcaaaccaaaaagggttcttacttctgaagctgaaaagatcgctcctcctgtactcattgcactgtttccc
1 M M L V K P T K G L L L A K A E K I A P P V L I A L F P
7272 ataccatgtctgaaagtattgcgaatgtttgtcttga 7234
29 I P C L K V L R M F C S *
1820RF065
3582 atgaatgctatctgtatcacataaataatgcgatcaaaacatttttgagcgggtgtaatggtagtatctaccccaagccgt
1 M N A I C I T I N N A I K T F L S G C N G S I S T P S R
3498 cacaactagcagcggaacataaacaggatctcttaa 3460
29 H K T S K R N I N R I S *
1820RF066
4234 atgtggctactctttttgtgtttcacagaattatgtttcacgtgaaacagtttttatgggtataatagaatcaaaaggaggtgg
1 M W L L F F V F H R I M F H V K Q F L W Y N R I K R R W
4318 agattatggaatcaagaacatgaatcaatttaa 4353
29 R L W K L K N M N Q F *
1820RF067
13882 atgatacctgcttttagcttcaaaactactaaagtcgatttcttctgttaatttggaattgtaaaacctaacacaaaatcgata
1 M I P A L A L K L L K S I S L L N L A I V K P N T K S I
13798 atcattgcaaccattaacctataatcaaacctaa 13763
29 I I A T I N H I I K P *
1820RF068
7267 atgtctgaaagtattgcgaatgttttctcttgagcaatcaaggagttttgttctcttgcagtaatgcagaagcatagtcaga
1 M S E S I A N V L L L S N Q G V F V S L H E C R S I V R
7183 ttttaactcctacatcgtaggatcattatcgattaa 7148
29 F N S Y I V R I I I D *
1820RF069
5027 gtggaacaatgtttttacatcgggaaacttctgttttaaatccctgttaacagactcgtaggggtgaaacttatgttctctgtgc
1 V E Q C F Y I G N F L F K Y P C N R L V R V E L M F L C
4943 aatgtcaacaaaatttcttcaatcggttcgacctaa 4908
29 N V N K N F F N R S T *
1820RF070
1031 gtgatgggttcggctccacaaaaccagaaacttctgtctgagtaaaactagaatatctttcaattctagtagtctcgccaattgcgt
1 V M V R L H Q N Q K L R L S K L E Y L S I L V L R Q L R
947 tacgtaacggaatgaagccccgtttgtggcattga 912
29 Y V T E L K P R L W H *
1820RF071
11741 atggttttgcattatgggtgccacaaagcgctcaagtggttaagggaattttctttaaatactcgcaattacaatcggtttg
1 M V L H Y G C H K A L K V V K E P S L M I L A I T I V L
11825 actttgattgttggctcgaactgtactttaa 11857
29 T L I C L F V T V L *
1820RF072
11723 atgtttacattaaatgcgcgtcattgtttcaaaactttatgtcgtttctcccgatcccaagaaagtaactacaggtacatcacgt
1 M F T L N A V I V S N F N V V S P D P K K V T T G T S R
11639 ttcaattcaatgggttagcaaacgataa 11610
29 F N S M V L A K R *
1820RF073
2876 gtgaagccgctttgtatgctttacgttaagctttatcaaaccttaagacaaaataggaaccattgtttgaaagtgtattt
1 V K P P L Y A L R K S L S N P K D K I G N H C L K V D F
2792 ccatgtgtagcttttagccaatctttgtaa 2763
29 P C V A F S Q S L *
1820RF074
8923 gtgattgataaattttgtttcaaatctgtcgttttctcgtcataaaacggataatcaaaatcaaacattgttttcggcc
1 V I D K F C P K F C S P C F V I K R I I K I K Q L F S A
8839 aacttcaatcggttttctgagataa 8813
29 N F N T F F S R *
1820RF075
7463 gtgttacattatcgggaatttttgcataatctgccactttacgtccaagaggttcaaacggttttcttttcagaaacatagt
1 V L H Y L E Y F R Y L P L Y L P R G S N R F L F Q K H S
7379 tgtttacttgggtcctgtcctccatga 7353
29 C L L V V L L P *
1820RF076
2426 atgagtgtggagaatgttcgatcttcttttctttacaccatttgaaaccatttttgaataaccatgaagcataaactct

328

1 M S V E N V R S S F A S L H H L K P F L N N H E S I N S
2342 ccgtcaaatttttcgttggtgaaataa 2316
29 P S N F S L W K *
1820R7077
11858 atgaaggaacgtatgttctgttctagaggtagaggggttacatttgaaaattgtctattctctaataatctctcaagcaatta
1 M K E R M L L L L E V E G L H L K I V Y S L I S L K Q L
11942 tcaaaacagcttttcccgatgtaa 11965
29 S K Q L F P M *
1820R7078
7671 gtgcctacaataatttggttcttttaatttaataatgaaattccatgcttttctgtttgtaagtttggtgtagctactcgattgctc
1 V P T I F G S F N L M K F H A F L V C K F G V A T R L L
7587 tttgtgccatacattgagaagtaa 7564
29 F V P Y I E K *
1820R7079
7488 gtgaagataagtttgatccaagctgtgttacattatctggaatattttcgatatctgccactttacctgccaagaggttcaaa
1 V K D K F D P S C V T L S G I F S I S A T L P A K R P K
7404 ccgttttcttttccagaaacatag 7381
29 P F S F S E T *
1820R7080
4473 gtgtgctatttgctgatgtcaaaagcttcagttgttgcctcgtagctttctcgcaatgcttcaagatgttctacaatctttgatc
1 V C Y L L M S K L Q L L L R S L L A M L Q D V L Q S L I
4389 ttgcttcaccgtctgtga 4372
29 L L H R L *

Table 24

Sequence similarities phage 182 and public databases

Phage: 182

Database: nr

Query= sid|110156|lan|182ORF001 Phage 182 ORF|5966-7780|2
(604 letters)

gi 138124 sp P07534 VG9_BPPZA TAIL PROTEIN (LATE PROTEIN GP9) >...	384	e-105
gi 138123 sp P04331 VG9_BPPH2 TAIL PROTEIN (LATE PROTEIN GP9) >...	374	e-103
gi 1429238 gnl PID e1173412 (X99260) tail protein [Bacteriophag...	346	3e-94
gi 215339 (M12456) p9 tail protein [Bacteriophage phi-29] >gi 2...	208	8e-53
gi 1181970 gnl PID e221269 (Z47794) tail protein [Bacteriophage...	62	8e-09
gi 1181968 gnl PID e221267 (Z47794) tail protein [Bacteriophage...	56	6e-07
gi 2500030 sp Q59968 CARA_SULSO CARBAMOYL-PHOSPHATE SYNTHASE SM...	49	8e-05

Query= sid|110157|lan|182ORF002 Phage 182 ORF|2152-3873|1
(573 letters)

gi 118848 sp P19894 DPOL_BPM2 DNA POLYMERASE >gi 76896 pir JQ0...	665	0.0
gi 1429230 gnl PID e1173404 (X99260) DNA polymerase [Bacterioph...	657	0.0
gi 118849 sp P03680 DPOL_BPPH2 DNA POLYMERASE (EARLY PROTEIN GP...	654	0.0
gi 118851 sp P06950 DPOL_BPPZA DNA POLYMERASE (EARLY PROTEIN GP...	654	0.0
gi 15732 (X53371) DNA polymerase (AA 1-575) [Bacteriophage phi-29]	651	0.0
gi 15734 (X53370) DNA polymerase (AA 1-575) [Bacteriophage phi-29]	651	0.0
gi 1572479 gnl PID e242301 (X96987) DNA polymerase [Bacterioph...	565	e-160
gi 1072656 pir S51275 DNA polymerase - phage CP-1 >gi 836593 g...	301	1e-80
gi 118847 sp P22374 DPOM_ASCIM PROBABLE DNA POLYMERASE >gi 8385...	71	3e-11
gi 461962 sp P33537 DPOM_NEUCR PROBABLE DNA POLYMERASE >gi 2833...	65	1e-09
gi 461963 sp P33538 DPOM_NEUTIN PROBABLE DNA POLYMERASE >gi 1018...	62	1e-08
gi 1084487 pir S41618 DNA polymerase - slime mold (Physarum po...	61	3e-08
gi 2435429 (AF012250) unassigned reading frame (possible DNA po...	61	3e-08
gi 578157 gnl PID e246743 (X52106) DNA polymerase (Neurospora i...	59	1e-07
gi 2147969 pir S72369 probable DNA-polymerase - Gelasinospora ...	58	2e-07
gi 2147968 pir S62752 probable DNA-polymerase - Gelasinospora ...	58	2e-07
gi 3511140 (AF061244) B type DNA polymerase [Agrocye aegerita]	57	3e-07
gi 118850 sp P10479 DPOL_BPPRD DNA POLYMERASE (PROTEIN P1) >gi ...	56	6e-07
gi 578144 (X63909) putative DNA-polymerase, B-type [Morchella c...	47	3e-04
gi 232013 sp P30322 DPOM_AGABT PROBABLE DNA POLYMERASE >gi 3208...	46	6e-04

Query= sid|110159|lan|182ORF004 Phage 182 ORF|4626-5954|3
(442 letters)

gi 138117 sp P13849 VG8_BPPH2 MAJOR HEAD PROTEIN (LATE PROTEIN ...	309	2e-83
gi 138118 sp P07531 VG8_BPPZA MAJOR HEAD PROTEIN (LATE PROTEIN ...	305	3e-82
gi 1429236 gnl PID e1173410 (X99260) major head protein [Bacter...	300	1e-80
gi 1181958 gnl PID e221257 (Z47794) major head protein [Bacteri...	152	6e-36

Query= sid|110160|lan|182ORF005 Phage 182 ORF|12651-13700|3
(349 letters)

gi 137932 sp P15132 VG13_BPPH2 MORPHOGENESIS PROTEIN 1 (LATE PR...	52	8e-06
gi 1429242 gnl PID e1173416 (X99260) morphogenesis protein [Bac...	48	7e-05
gi 137933 sp P07538 VG13_BPPZA MORPHOGENESIS PROTEIN 1 (LATE PR...	47	2e-04

Query= sid|110161|lan|182ORF006 Phage 182 ORF|14995-16026|1
(343 letters)

gi 137944 sp P11014 VG16_BPPH2 ENCAPSIDATION PROTEIN (LATE PROT...	402	e-111
gi 137945 sp P07541 VG16_BPPZA ENCAPSIDATION PROTEIN (LATE PROT...	402	e-111
gi 1429245 gnl PID e1173419 (X99260) encapsidation protein [Bac...	381	e-105
gi 1181972 gnl PID e221271 (Z47794) encapsidation protein [Bact...	159	2e-38

Query= sid|110162|lan|182ORF007 Phage 182 ORF|7795-8775|1
(326 letters)

gi 1429239 gnl PID e1173413 (X99260) upper collar protein [Bact...	271	5e-72
gi 137915 sp P07535 VG10_BPPZA UPPER COLLAR PROTEIN (CONNECTOR ...	256	1e-67
gi 137914 sp P04332 VG10_BPPH2 UPPER COLLAR PROTEIN (CONNECTOR ...	256	2e-67
gi 1181960 gnl PID e221259 (Z47794) connector protein [Bacterio...	148	6e-35

Query= sid|110163|lan|182ORF008 Phage 182 ORF|14105-14983|2
(292 letters)

gi 4210750 gnl PID e1374037 (AJ132604) LysL protein [Lactococcu...	139	2e-32
gi 462559 sp P34020 LYC_CLOAB AUTOLYTIC LYSOZYME (1,4-BETA-N-AC...	75	8e-13
gi 2327014 (U82823) putative lysozyme [Saccharopolyspora erythr...	64	2e-09
gi 126652 sp P25310 LYCM_STRGL LYSOZYME M1 PRECURSOR (1,4-BETA-...	60	2e-08
gi 127789 sp P19386 LYCA_BPCP9 LYSOZYME (ENDOLYSIN) (MURAMIDASE...	60	2e-08
gi 67761 pir MUBPCP N-acetylmuramoyl-L-alanine amidase (EC 3.5...	59	3e-08
gi 4105636 (AF049087) lys [Leuconostoc oenos bacteriophage 10MC]	59	3e-08
gi 623084 (L02496) muramidase; muramidase [Bacteriophage LL-H]	57	1e-07
gi 127787 sp P15057 LYCA_BPCP1 LYSOZYME (ENDOLYSIN) (MURAMIDASE...	57	2e-07
gi 126597 sp P00721 LYCH_CHASP N,O-DIACETLYLMURAMIDASE (LYSOZYME...	57	2e-07
gi 127788 sp P19385 LYCA_BPCP7 LYSOZYME (ENDOLYSIN) (MURAMIDASE...	57	2e-07
gi 67762 pir MUBPC7 N-acetylmuramoyl-L-alanine amidase (EC 3.5...	56	3e-07
gi 3025168 sp P76421 YEGX_ECOLI HYPOTHETICAL 32.0 KD PROTEIN IN...	53	2e-06
gi 4204413 (AF047001) Lys44 [Oenococcus oeni temperate bacterio...	53	3e-06
gi 2116978 gnl PID d1020940 (D88151) cortical fragment-lytic en...	52	5e-06
gi 2392844 (AF011378) lysin [Bacteriophage sk1]	48	8e-05

Query= sid|110164|lan|182ORF009 Phage 182 ORF|8765-9601|2
(278 letters)

gi 1429240 gnl PID e1173414 (X99260) lower collar protein [Bact...	180	1e-44
gi 137921 sp P04333 VG11_BPPH2 LOWER COLLAR PROTEIN (LATE PROTE...	171	5e-42
gi 215341 (M12456) p11 lower collar protein [Bacteriophage phi-29]	98	9e-20
gi 224162 prf 1011232B protein p11, lower collar [Bacteriophage...	97	1e-19
gi 535260 (Z30339) STARP antigen [Plasmodium reichenowi]	50	1e-05
gi 4049753 (AF063866) ORF MSV230 hypothetical protein [Melanopl...	49	4e-05
gi 2131557 pir S70306 hypothetical protein YEL077c - yeast (Sa...	48	5e-05
gi 131782 sp P12753 RAS0_YEAST DNA REPAIR PROTEIN RAD50 (153 KD...	48	7e-05
gi 2131309 pir S70305 hypothetical protein YBL113c - yeast (Sa...	47	2e-04
gi 499325 (Z26314) STARP antigen [Plasmodium falciparum]	46	3e-04
gi 3845171 (AE001391) ribosome releasing factor (OO, TP) [Plasm...	46	3e-04
gi 731903 sp P40434 YIR7_YEAST HYPOTHETICAL 197.5 KD PROTEIN IN...	45	5e-04
gi 1632829 gnl PID e276379 (Y08924) AARP2 protein [Plasmodium f...	45	5e-04
gi 1176490 sp P40889 YJW5_YEAST HYPOTHETICAL 197.6 KD PROTEIN I...	45	5e-04
gi 1077300 pir S51848 hypothetical protein HRD1054 - yeast (Sa...	45	5e-04
gi 2425143 (AF020407) Wima [Dictyostelium discoideum]	45	6e-04
gi 1181961 gnl PID e221260 (Z47794) collar protein [Bacterioph...	45	6e-04
gi 2132657 pir S64819 probable membrane protein YLL067c - yeas...	45	8e-04
gi 2133041 pir S65341 probable membrane protein YPR204w - yeas...	45	8e-04
gi 730275 sp P39793 PBPA_BACSU PENICILLIN-BINDING PROTEINS 1A/1...	45	8e-04

Query= sid|110165|lan|182ORF010 Phage 182 ORF|13110-2155|2
(281 letters)

gi 135604 sp P06812 TERM_BPNF DNA TERMINAL PROTEIN >gi 75815 pi...	69	3e-11
gi 1572478 gnl PID e242334 (X96987) terminal protein [Bacteriop...	65	3e-10
gi 1429231 gnl PID e1173405 (X99260) terminal protein [Bacterio...	64	1e-09

Query= sid|110166|lan|182ORF011 Phage 182 ORF|9607-10158|1
(183 letters)

gi 137928 sp P07537 VG12_BPPZA PRE-NECK APPENDAGE PROTEIN (LATE...	51	6e-06
gi 1429241 gnl PID e1173415 (X99260) pre-neck appendage protein...	51	6e-06
gi 137927 sp P20345 VG12_BPPH2 PRE-NECK APPENDAGE PROTEIN (LATE...	50	1e-05

Query= sid|110169|lan|182ORF014 Phage 182 ORF|13716-14108|3
(130 letters)

gi 137936 sp P11188 VG14_BPPH2 LYSIS PROTEIN (LATE PROTEIN GP14...	97	6e-20
gi 137938 sp P07539 VG14_BPPZA LYSIS PROTEIN (LATE PROTEIN GP14...	96	8e-20
gi 1429243 gnl PID e1173417 (X99260) lysis protein [Bacterioph...	96	8e-20
gi 215332 (M14782) lysis protein [Bacteriophage phi-29]	94	5e-19

Query= sid|110170|lan|182ORF015 Phage 182 ORF|854-1225|2
(123 letters)

331

gi|15670 (V01155) reading frame 10 (may be gene 4) (Bacterioph... 70 5e-12
 gi|138072|sp|P06953|VG5A_BPPZA EARLY PROTEIN GP5A >gi|75836|pir... 69 7e-12

Query= sid|110174|lan|182ORF019 Phage 182 ORF|4323-4613|3
 (96 letters)

gi|1429235|gnl|PID|e1173409 (X99260) head morphogenesis protein... 61 2e-09
 gi|138111|sp|P13848|VG7_BPPH2 HEAD MORPHOGENESIS PROTEIN (LATE ... 57 3e-08
 gi|138112|sp|P07533|VG7_BPPZA HEAD MORPHOGENESIS PROTEIN (LATE ... 54 1e-07

Query= sid|110180|lan|182ORF025 Phage 182 ORF|548-814|2
 (88 letters)

gi|138099|sp|P06955|VG6_BPPZA EARLY PROTEIN GP6 >gi|75841|pir|... 55 7e-08
 gi|138098|sp|P03685|VG6_BPPH2 EARLY PROTEIN GP6 >gi|75840|pir|... 54 2e-07
 gi|1429234|gnl|PID|e1173408 (X99260) gene 6 product (Bacterioph... 54 2e-07

Table 25

Homologies between 182 ORFs and proteins in public databases

Phage: 182

Database: Swissprot

Query= sid|110156|lan|182ORF001 Phage 182 ORF|5966-7780|2
(604 letters)

gi 138124 sp P07534 VG9_BPPZA TAIL PROTEIN (LATE PROTEIN GP9)	384	e-106
gi 138123 sp P04331 VG9_BPPH2 TAIL PROTEIN (LATE PROTEIN GP9)	374	e-103
gi 2500030 sp Q59968 CARA_SULSO CARBAMOYL-PHOSPHATE SYNTHASE SM...	49	2e-05

Query= sid|110157|lan|182ORF002 Phage 182 ORF|2152-3873|1
(573 letters)

gi 118848 sp P19894 DPOL_BPM2 DNA POLYMERASE	665	0.0
gi 118849 sp P03680 DPOL_BPPH2 DNA POLYMERASE (EARLY PROTEIN GP2)	654	0.0
gi 118851 sp P06950 DPOL_BPPZA DNA POLYMERASE (EARLY PROTEIN GP2)	654	0.0
gi 118847 sp P22374 DPOM_ASCIM PROBABLE DNA POLYMERASE	71	7e-12
gi 461962 sp P33537 DPOM_NEUCR PROBABLE DNA POLYMERASE	65	3e-10
gi 461963 sp P33538 DPOM_NEUIN PROBABLE DNA POLYMERASE	62	3e-09
gi 118850 sp P10479 DPOL_BPPRD DNA POLYMERASE (PROTEIN P1)	56	2e-07
gi 232013 sp P30322 DPOM_AGABT PROBABLE DNA POLYMERASE	46	2e-04
gi 118887 sp P10582 DPOM_MAIZE DNA POLYMERASE (S-1 DNA ORF 3)	46	2e-04

Query= sid|110159|lan|182ORF004 Phage 182 ORF|4626-5954|3
(442 letters)

gi 138117 sp P13849 VG8_BPPH2 MAJOR HEAD PROTEIN (LATE PROTEIN ...	309	6e-84
gi 138118 sp P07531 VG8_BPPZA MAJOR HEAD PROTEIN (LATE PROTEIN ...	305	7e-83

Query= sid|110160|lan|182ORF005 Phage 182 ORF|12651-13700|3
(349 letters)

gi 137932 sp P15132 VG13_BPPH2 MORPHOGENESIS PROTEIN 1 (LATE PR...	52	2e-06
gi 137933 sp P07538 VG13_BPPZA MORPHOGENESIS PROTEIN 1 (LATE PR...	47	6e-05

Query= sid|110161|lan|182ORF006 Phage 182 ORF|14995-16026|1
(343 letters)

gi 137945 sp P07541 VG16_BPPZA ENCAPSIDATION PROTEIN (LATE PROT...	402	e-112
gi 137944 sp P11014 VG16_BPPH2 ENCAPSIDATION PROTEIN (LATE PROT...	402	e-112

Query= sid|110162|lan|182ORF007 Phage 182 ORF|7795-8775|1
(326 letters)

gi 137915 sp P07535 VG10_BPPZA UPPER COLLAR PROTEIN (CONNECTOR ...	256	3e-68
gi 137914 sp P04332 VG10_BPPH2 UPPER COLLAR PROTEIN (CONNECTOR ...	256	5e-68

Query= sid|110163|lan|182ORF008 Phage 182 ORF|14105-14983|2
(292 letters)

gi 462559 sp P34020 LYC_CLOAB AUTOLYTIC LYSOZYME (1,4-BETA-N-AC...	75	2e-13
gi 126652 sp P25310 LYCM_STRGL LYSOZYME M1 PRECURSOR (1,4-BETA-...	60	5e-09
gi 127789 sp P19386 LYCA_BPCP9 LYSOZYME (ENDOLYSIN) (MURAMIDASE...	60	5e-09
gi 127787 sp P15057 LYCA_BPCP1 LYSOZYME (ENDOLYSIN) (MURAMIDASE...	57	4e-08
gi 126597 sp P00721 LYCH_CHASP N,O-DIACETYL MURAMIDASE (LYSOZYME...	57	4e-08
gi 127788 sp P19385 LYCA_BPCP7 LYSOZYME (ENDOLYSIN) (MURAMIDASE...	57	5e-08
gi 3025168 sp P76421 YEGX_ECOLI HYPOTHETICAL 32.0 KD PROTEIN IN...	53	5e-07

Query= sid|110164|lan|182ORF009 Phage 182 ORF|8765-9601|2
(278 letters)

gi 137921 sp P04333 VG11_BPPH2 LOWER COLLAR PROTEIN (LATE PROTE...	171	1e-42
gi 131782 sp P12753 RA50_YEAST DNA REPAIR PROTEIN RAD50 (153 KD...	48	2e-05
gi 1176490 sp P40889 YJW5_YEAST HYPOTHETICAL 197.6 KD PROTEIN I...	45	1e-04
gi 731903 sp P40434 YIR7_YEAST HYPOTHETICAL 197.5 KD PROTEIN IN...	45	1e-04
gi 730275 sp P39793 PBPA_BACSU PENICILLIN-BINDING PROTEINS 1A/1...	45	2e-04
gi 1168610 sp P41696 AZF1_YEAST ASPARAGINE-RICH ZINC FINGER PRO...	44	3e-04

333

gi 731587 sp P38900 YH19_YEAST HYPOTHETICAL 70.1 KD PROTEIN IN ...	44	3e-04
Query= sid 110165 lan 182ORF010 Phage 182 ORF 1310-2155 2 (281 letters)		
gi 135604 sp P06812 TERM_BPNF DNA TERMINAL PROTEIN	69	8e-12
Query= sid 110166 lan 182ORF011 Phage 182 ORF 9607-10158 1 (183 letters)		
gi 137928 sp P07537 VG12_BPPZA PRE-NECK APPENDAGE PROTEIN (LATE...	51	2e-06
gi 137927 sp P20345 VG12_BPPH2 PRE-NECK APPENDAGE PROTEIN (LATE...	50	3e-06
Query= sid 110169 lan 182ORF014 Phage 182 ORF 13716-14108 3 (130 letters)		
gi 137936 sp P11188 VG14_BPPH2 LYSIS PROTEIN (LATE PROTEIN GP14)	97	2e-20
gi 137938 sp P07539 VG14_BPPZA LYSIS PROTEIN (LATE PROTEIN GP14)	96	2e-20
Query= sid 110170 lan 182ORF015 Phage 182 ORF 854-1225 2 (123 letters)		
gi 138072 sp P06953 VG5A_BPPZA EARLY PROTEIN GP5A	69	2e-12
Query= sid 110174 lan 182ORF019 Phage 182 ORF 4323-4613 3 (96 letters)		
gi 138111 sp P13848 VG7_BPPH2 HEAD MORPHOGENESIS PROTEIN (LATE ...	57	9e-09
gi 138112 sp P07533 VG7_BPPZA HEAD MORPHOGENESIS PROTEIN (LATE ...	54	4e-08
Query= sid 110180 lan 182ORF025 Phage 182 ORF 548-814 2 (88 letters)		
gi 138099 sp P06955 VG6_BPPZA EARLY PROTEIN GP6	55	2e-08
gi 138098 sp P03685 VG6_BPPH2 EARLY PROTEIN GP6	54	5e-08

334

BLASTP 2.0.8 (Jan-05-1999)

Query= sid|110156|lan|182ORF001 Phage 182 ORF|5966-7780|2
(604 letters)

>gi|138124|sp|P07534|VG9_BPPZA_TAIL_PROTEIN (LATE PROTEIN GP9)
>gi|75849|pir|WMBP92 gene 9 protein - phage PZA
>gi|216058 (M11813) tail protein (Bacteriophage PZA)
Length = 599

Score = 384 bits (975), Expect = e-105
Identities = 231/610 (37%), Positives = 344/610 (55%), Gaps = 36/610 (5%)

Query: 6 TNVKKLANVFPDNTYTHTRWFKTQQEQESYFNSFPVLNENRDCSYQRDTQLGGVPRVDKH 65
TNV++LA+VFP N Y +TRWF + Q ++FNS + E ++Q + V
Sbjct: 9 TNVRILADVPPSNDYKVRWFTSSSNQYNWFNSKTRVYEMSKVTFQGFRENKSYISVSLR 68

Query: 66 KDALYACNYLIFKNEETYPKQYAFVTDIEYKNDNTSFVTFEIDVLQTYRFDIGIRESF 125
D LY +Y++F+N + Y +KW YAFVT++EYKN T++V FEIDVLQT+ F+I +ESF
Sbjct: 69 LLLYNASYIMFQAD-YGNKWFYAFVTELEYKNGVTTYVHFIDVLQTMFNIKFQESF 127

Query: 126 IAKEHPQLYYSNGIPFINTIEESLDYGREYTTTNTVTFHPNDGVNFLVILTSEAM--PVG 183
I +SH +L+ +G P INTI+E L+YG EY +V P D + FLV+++ M G
Sbjct: 128 IYREHVKLWNDGCTPTINTIDEGLNYGSEYDIVSVENHRPYDDMMFLVVISKSIMHGCTAG 187

Query: 184 DKEDKSG---GSIVGGSPSPFSYLLPINSSGEVYKPN-GAGNANFGEYMAFLT---TKEP 236
+ E + S+ G P P YY+ P G+V K G NAN + LT +++
Sbjct: 188 EAESRLNDINASLNGMPQLCYIHPYKDGKVPKTFIGDNNANLSPIVNMLTNIPSQKS 247

Query: 237 PLNKIVGMVYVTSYTGIFPVDHANKTVRYNAGGSYKIMLPTYASDPTGTMKTFAFFCVKE 296
+N IV MYVT Y G+ + +K ++ + + + A D G + T VK+
Sbjct: 248 AVNNIVNMVYTDYIGLKLDYKNGDKELKLDKDMFEQAGI---ADKKGNVDTIF---VKK 301

Query: 297 ARTFVPKRIDLVGNVYNYFREAFPNVSKSLFMYPCYCLIEITDTKGHVMTLRPEYLTGG 356
+ ID G+ + F + +ESKL MYPYC+ E+TD KG+ M L+ EY+
Sbjct: 302 IPDYETLEID-TGDKWGGPTKD-----QESKLMMYPYCVTEVTDKGNHNMNLKTEYIDNN 355

Query: 357 KLSVYVKGSLGISNKMIEPIDYDVSNSTI----ITNLSDKMLIDNDPNDVGKSDYASA 412
KL + V+SLG+SNKV DY+ S +T D LI+N+PND+ + +DY SA
Sbjct: 356 KKKIQVRGSLGVSNKVAYSIQDYNAGGSLSGGDRLTASLDTSLINNNPNDIAIINDYLSA 415

Query: 413 FMQGNKNSLIAEQNIRMTFRHGMGNSAMSTGGAIFSAASNNPFVGLTNIMGAGQQVNN 472
++QGNKNSL Q+ +I GM +S G ++ +PP +++ G N
Sbjct: 416 YLQGNKNSLENQKSSILFNGIVGMLGGGVSA-----ASAVGRSPFGLASSVTGTMSTAGN 471

Query: 473 YVSEKENGLNLLAGKVADIENIPDNVTQLGSNLSFTTGN-PQNYQLRFKQIKYEYATRL 531
V + + L K ADI NIP +T++G N +F GN ++ Y ++ KQ+K EY L
Sbjct: 472 AVLD----MQALQAKQADIANIPQLTKMGNTAFDYGNVGRGVYVIK-KQLKAEYRRSL 526

Query: 532 DRYFSMYGTSKNRVATPNLQTRKAWNFIKLKEPNIVGTMSNDVLTRVKQIFSAGVTLWHT 591
+F YG K NRV PNL+TRKA+N+I+ K+ I G ++N+ L ++ IF G+TLWHT
Sbjct: 527 SSFPHKYGYKINRVKPNLRTKAYNYIQTDCDFISGDINNNDLQEI RTIFDNGITLWHT 586

Query: 592 NDVLNYYQDN 601
+D+ NY+ +N
Sbjct: 587 DDIGNYSVEN 596

Query= sid|110157|lan|182ORF002 Phage 182 ORF|2152-3873|1
(573 letters)

>gi|118848|sp|P19894|DPOL_BPM2 DNA POLYMERASE >gi|76896|pir|JQ0161
DNA-directed DNA polymerase (EC 2.7.7.7) - phage M2
>gi|215509 (M33144) DNA polymerase (Bacteriophage M2)
Length = 572

Score = 665 bits (1697), Expect = 0.0
Identities = 327/589 (55%), Positives = 420/589 (70%), Gaps = 38/589 (6%)

Query: 3 KKYTGDFETTTDLNDCRVWSWGVCDIDNVNMTFGLSIDSPFENCKMOGSTDIYFHNEKF 62
K ++ DFETTT L+DCRVW++G +I N+DN G +D F +W M+ D+YFHN KP

335

Sbjct: 4 KMFSCDFETTTKLDRCRVNAYGYMEIGNLDNYKIGNSLDEFMQWV-MEIQADLYFHNLF 62

Query: 63 DGEFMLSFLFKNGFKWCKEAKEDRTFSTLISNMGOYALEICWEVNYXXXXXXXXXXXX 122
DG F++WL ++GFKW E + T++IS MGQWY ++IC+

Sbjct: 63 DGAFIVNWLEQHGFKWSNEGLPN-TYNTIISKMGQWYMIDICFGYK-----GKRKL 112

Query: 123 XXIIYDSLKKYFPFVKQIAEAFNFPPIKKGEIDYTKERPIGYKPTKDEWEYLKNDIQIMAM 182
+IYDSLKK FFPVK+IA+ F P+ KG+IDY ERP+G++ T +E+EY+KNDI+I+A

Sbjct: 113 HTVIYDSLKKLFPFVKKIAKDFQLPLKGDIDYHTERPVGHEITPEEYIYKNDIEIAR 172

Query: 183 ALKIQFDQGLTRMTRGSDALGDYKDWLKATHGKSTFKQWFPILSLGFDKDLRKAYKGGFT 242
AL IQF QGL RMT GSD+L +KD L F + FP LSL DK++RKAY+GGFT

Sbjct: 173 ALDIQFKQGLDRMTAGSDSLKGFKDILST----KCFNKVFPKLSLPMDEIRKAYRGGFT 228

Query: 243 WVNKVFQKGEIGDGVFDVNSLYPSQMYVRPLPYGTPLFYEGEYKPNNDYPLYIQNIKVR 302
W+N ++ KEIG+G+VFDVNSLYPSQMY RPLPYG P+ ++G+Y+ + YPLYIQ I+

Sbjct: 229 WLNDKYKEKEIGEMVFDVNSLYPSQMYSRPLPYGAPIVFQKYEKDEQYPLYIQRIRFE 288

Query: 303 FRLKEGYIPTIQVKQSSLFIQNEYLESSVNLGVDELIDLTLTNVDELFFEHYDILEIH 362
F LKEGYIPTIQ+K++ F NEYL++S GV E ++L LTNVDELEL EHY++ +

Sbjct: 289 FELKEGYIPTIQIKKNPFFKNGEYLNKNS----GV-EPVELYLTNVDELELIQEHYELNYE 343

Query: 363 YTYGYMFKASCDFMGWIDKWIEVKNTEGARKANAKGMLNSLYGKFGTNPDTGKVPYM 422
Y G+ F+ +FK +IDKW VK EGA+K AK MLNSLYGKF +NPD+TGKVPY+

Sbjct: 344 YIDGPKFREKTGLFKDFIDKWTYVKTHEEGAKQLAKLMLNSLYGKFASNPVDTGKVPYL 403

Query: 423 GEDGIVRLTLGEEELRDPVYVPLASFVTAWGRYTTITTAQKCFDRRIYCDTDSIHLVGTE 482
+DG + +G+EE +DPVY P+ F+TAW R+TTIT AQ C+DRIYCDTDSIHL GTE

Sbjct: 404 KDDGSLGFRVGDDEYKDPVYTPMGVFTITAWARFTTITAAQACYDRIYCDTDSIHLTGTE 463

Query: 483 VPEAIDHLVDPKKLGWGHSTFQRAKFIQKT-----YVEEIDGEL----- 524
VPE I +VDPKKLGYW HESTF+RAK++RQKT YV+E+DG+L

Sbjct: 464 VPEIKDIVDPPKLGYNHSTFQRAKYLQKTYIQDIYVKEVDGKLKESPDDEATTTKF 523

Query: 525 NVKACAGMPDRIKEIVTFDNFVGFSSYGLKLPKRTQGGVVLVDVTMTTIK 573
+VKACAGM D IK+ VTFDNF VGFSS GK P + GGVVLVD++PTIK

Sbjct: 524 SVKACAGMTDTIKKKVTFDNFAVGFSMKGKPKPVQVNGGVVLVDVSVFTIK 572

Query= sid|110159|lan|182ORF004 Phage 182 ORF|4626-5954|3
(442 letters)

>gi|138117|sp|P13849|VG8_BPPH2 MAJOR HEAD PROTEIN (LATE PROTEIN GP8)
>gi|75845|pir||WMBP89 gene 8 protein - phage phi-29
>gi|215325 (M14782) major head protein [Bacteriophage
phi-29] >gi|225362|prf||1301270B gene 8 [Bacillus sp.]
Length = 448

Score = 309 bits (783), Expect = 2e-83
Identities = 176/440 (40%), Positives = 250/440 (56%), Gaps = 27/440 (6%)

Query: 4 KITEQDVLRAITVETPVQLMTAIYNSSSSLFQANVMPNADNIEAVGAGITRLDVVKNF 63
+IT DV + + ++ AI NS F++ VP+ A+N+ VGAGI V+N+P

Sbjct: 2 RITFNDVKTSLGITESYDIVNAIRNSQGDNFKSYVPLATANNAEAVGAGILINQTVQNDP 61

Query: 64 ISTLVDRIGKVIRYKSWRNPLKMFKKGMPLGRTIEEIFVDIAQEHKFNPDSESVTGVEK 123
I++LVDRIG VVIR S NPLK FKKG +PLGRTIEEI+ DI +E +++ +E+ VP+

Sbjct: 62 ITS LVDRIGLVIRQVSLNNPLKFKKGQIPLGRTIEEIIYTDITKEKQYDABEAEQKVFE 121

Query: 124 QEVPDVKTLFHEINREGYYKQTIQEAMLEKAPTSDNFNFSFVAGVMNLYTGDVSEFEY 183
+E+P+VKTLPHE NR+G+Y QTIQ+ L+ AF SW NF SFV+ ++NA+Y EV E+EY

Sbjct: 122 REMPNVKTLPHERNRQGPYHQTIQDDSLKTAFVSWGNFESFVSSIINAIYNSAEVDEY 181

Query: 184 TKLLIANYQEKELFKEIEIGEITESNA--KEFIRKIKSTSNKLEFM--SSAYNAQGVKTS 239
KLL+ NY K LF ++I E T S EF++K+++T+ KL S +N+ V+T

Sbjct: 182 MKLLVDNYSKGLFTTVKIDEPTSSTGALTEFVKMRATARKLTLPQGSRDWNSMAVRTR 241

Query: 240 TSKSDQYKXXXXXXXXXXXXXXXXXFNMSKTDVGHKIVIDEFPKKEGEESNIVAVIV 299
+ D + FNM++TDF+G+ VID F S+ + AV+V

Sbjct: 242 SYMEDLHLIIDADLEAELDVDVLAKAFNMNRTDPLGNVTVIDGF-----ASTGLEAVLV 295

Query: 300 DSEWPMIYDKLYKTTSLYNPEGLYWNYWLHHQLYSTSQFGNAFAVFSATKPVTKVAPA 359
D +WPM+YD L+K ++ NP GLYWNY+ H Q S +F NAVAFAV VT+V +

Sbjct: 296 DKDWMFVYDNLHKMETVRNPRGLYWNYHHVWQTLVSVSRFANAVAFVSGDVPVAVTQVIVS 355

Query: 360 SATTSVVKSSKDIALTFTPEATNQGEVVSSAPALVKATVKQTAGKATAVTVEGLEV 419

336

+V +G + V ATN + V V G +T + G
 Sbjct: 356 PNIAAVKQGGQQQFT--AYVRATNAKDHKV-----VWSVEGGSTGTAI----TG 398

Query: 420 QSLVTFTAIGGQATVLVTV 439

L++ + Q TV TV

Sbjct: 399 DGLLSVSGNEDNQLTVKATV 418

Query= sid|110160|lan|182ORF005 Phage 182 ORF|12651-13700|3
 (349 letters)

>gi|137932|sp|P15132|VG13_BPPH2 MORPHOGENESIS PROTEIN 1 (LATE
 PROTEIN GP13) >gi|75858|pir|WMBP23 gene 13 protein -
 phage phi-29 >gi|215331 (M14782) morphogenesis protein
 [Bacteriophage phi-29] >gi|225368|prf|1301270H gene 13
 [Bacteriophage phi-29]
 Length = 365

Score = 51.5 bits (121), Expect = 8e-06

Identities = 44/166 (26%), Positives = 70/166 (41%), Gaps = 14/166 (8%)

Query: 6 NEQIARGQTIKILSKYGYNKNSQGVVLANLHWSA---GLNPNSNEXXXXXXXX-QWT 61
 +E Q I LS G+ K + G++ N+ ES GL N +E QWT

Sbjct: 12 SEMKVNAQYILNYLSSNGWTKQAICGLGNMQSESTINPGLWQNLDEGNTSLGFGLVQWT 71

Query: 62 PKSNIYRQAQICGLSNAKAETLEGQAEIIAQGDKTGQWMDNTPVSSAGYTNPQTLSPFKQ 121
 P SN A GL ++ II + + QW++ ++ Y K

Sbjct: 72 PASNYINWANSQGLPYKDMS--ELKRIWEVNNNAQWINLRDMTFKEY-----IKS 121

Query: 122 SANIDVATINFMCHWERPGKLHIEERLDLAQAYSKHIDGSGGGGVK 167
 + + F+ +ERP + ER D A+ + K++ G GGGG++

Sbjct: 122 TKTPRELAMIFLASIERPANPNQPERGDAEYWKYKNSGGGGGGLQ 167

Query= sid|110161|lan|182ORF006 Phage 182 ORF|14995-16026|1
 (343 letters)

>gi|137945|sp|P07541|VG16_BPPZA ENCAPSIDATION PROTEIN (LATE PROTEIN
 GP16) >gi|75861|pir|WMBP16 gene 16 protein - phage PZA
 >gi|216065 (M11813) morphogenesis protein C
 [Bacteriophage PZA]
 Length = 332

Score = 402 bits (1023), Expect = e-111

Identities = 186/332 (56%), Positives = 244/332 (73%), Gaps = 2/332 (0%)

Query: 11 EKNLYYNNPNNALGFNCLMLFVIGARGIGKTYGYKKFVNNRFIKHGEQFIYLRPFKTELKK 70
 +K+L+YNP L ++ ++ FVIGARGIGK+Y K + +NRFIK+GEQFIY+RR+K EL K

Sbjct: 2 DKSLFYNNPQKMLSYDRILNFVIGARGIGKSYAMKVYPINRFIKYGEQFIYVRRYKPELAK 61

Query: 71 IPQFFKTKAKEFPDHLKLEVKGEFYCDKLMGWAUPLSTWIEKSNEYPEVRTILPDEF 130
 + +F +A+EFFPDH+L VKG+ FY D KL GWA+PLS W EKS N YP V TI+PDEF+

Sbjct: 62 VSNYFNDVAQEFDPDHLVVKGRFVIDGKLAWAIPLSVWQSEKSNAYPNVSTIVDEFI 121

Query: 131 IEKSKITYLPNEAEALLMMETVFRRTNTRCVMLSNATSVVNPYFLYFNLQPDNLKRFN 190
 EK Y+PNE ALLN+M+TVFR R RC+ LSNA SVVNPYFL+FNL PD+NKRFN

Sbjct: 122 REKDNSNYIPNEVSALLNMDTVFRNRERVRCICLSNAVSVVNPYFLFNLVPOVNRFN 181

Query: 191 LYQDRGILIELCDSDKFAEVKRETPFGRLIRGTEYEDFSINNEFVNDSDTFIEKRSKSS 250
 +Y D LIE+ DS DF+ +R+T FGRLI GTEY + S++N+F+ DS FIEKRSK+S

Sbjct: 182 VYDD--ALIEIPDSLDFSSERRKTRFGRLIDGTEYGEMSLDNQFIGDSHVFIKRSKDSK 239

Query: 251 FLCAIAFEGKIFGYWIDAETGCVYVSVDYQFNTNHFYAMITKDHEENRLLMKNNRNNYYL 310
 F+ +I + G G W+D G +YV + P+T + Y +TT D EN +L+ N++NNY+L

Sbjct: 240 FVFSIVYNGFTLGWVDVNNQGLMYVDTAHPSTKGVYTLTTDDLNNMMLITNYKNNYHL 299

Query: 311 STVAKAFKNSYLRFDNIVIKNLHYDLFNKMKI 342
 +A AF N YLRFDN VI+N+ Y+LF KM+I

Sbjct: 300 RKLASAFMNGYLRFDNQVIRNIAYELFRKMRI 331

Query= sid|110162|lan|182ORF007 Phage 182 ORF|7795-8775|1
 (326 letters)

>gi|1429239|emb|CAA67658| (X99260) upper collar protein
 [Bacteriophage B103]

337

Length = 308

Score = 271 bits (685), Expect = 6e-72

Identities = 131/275 (47%), Positives = 187/275 (67%), Gaps = 5/275 (1%)

Query: 36 YYEHYRRQLTLLTFQLEWENLPKSIDPRYLEIALHTNGYLGFFKDPTLGFVMVCAEDG 95
 +Y HY + L L +QLFEWE LP S+DP YLE ++H GY+GF+KDP +G++ C GA G
 Sbjct: 22 WYHYHYQYLCSLAYQLFEWERLPPSVDPSPYLEKSIHQFGYGVGYKDPRIGYIACQGALSG 81

Query: 96 QIDHYHNPIFFTANEAMYHKRYPVLYRDDDDDKSKCIMLYNNDLKVPTLPSLHRFALDMA 155
 +DHY+ P F A+ Y + + Y D +K+ + +YNNDLK TLP+L FA D+A
 Sbjct: 82 TVDHYNLPRDFHASSVGYQNTFKLYNSDMKEKMGVAIYNNDLKCSTLPALEMPAQDLA 141

Query: 156 DINQISRVNRRQAQKTPVVIQTDEKKYFSLQAYNQIDENNQAVFVDKDMFEDESFNWQT 215
 ++ +I VN+ AQKTPV+I ++ SL YNQ + N +FV + ++ D + V++T
 Sbjct: 142 ELKEIIAVNQNAQKTPVLIANDNNQLSLKNIYNQYEGNAPVIFVHESLDLD-NLKVFKT 200

Query: 216 NAPYVVDKLRSELNEVNEVLTFLGINNANVDKTARVQTSVLSNNEQIESSGNILLKSR 275
 +APYVVDKL ++ N VMNEV+T+LGI NAN++K R+ TSEV SN+EQIESSGNI LK+R
 Sbjct: 201 DAPYVVDKLNAAQNAVNEVMTYLGIKNANLEKKERMVTSEVDSNDEQIESSGNIYLKAR 260

Query: 276 KEFCDRVNRVFGDELGKIDVKFRITDAVRQLQLAA 310
 +E C++++ ++G L VKFR D V Q++L A
 Sbjct: 261 QEACNKISELYGLNL----KVKFRYDIVEQMRLNA 291

Query= aid|110163|lan|182ORF008 Phage 182 ORF|14105-14983|2
 (292 letters)

>gi|4210750|emb|CAA10710| (AJ132604) LysL protein [Lactococcus
 lactis]
 Length = 235

Score = 139 bits (347), Expect = 2e-32

Identities = 85/210 (40%), Positives = 114/210 (53%), Gaps = 14/210 (6%)

Query: 2 MNGIDISSYQTGIDLSKVPCDFVNIKATGGTGYVNPDCDRAFOQALSIGKKIGVYHFAHE 61
 MNGIDISSYQ ++ VP DFV IKAT GT Y+NP + Q + K +G YHFA
 Sbjct: 1 MNGIDISSYQAEINAGIVPSDFVIKATECTNYINPTWEEQAGQVIQTNKLLGFYHFA- 59

Query: 62 RGLEGTPOQEAQFFLDNIKGYIGKAVLILDFEGS--NQKDVNNAKAFDYVYNTKGVKAW 119
 G P EA FF+ +K YIGKAVL+LDPE N A+ FL+ V KTG+
 Sbjct: 60 ---VGNPIABADFFISVVGNYIGKAVLVLDPEAGAINAWGNVGARQPLNRVKEKTGINPM 116

Query: 120 FYTYTANLNTTDFSSIAKGDYGLWVAEYGSNQPGYSQPAPPKTN-----FPIVACFPQ 174
 Y + ++S+I+ + LWVA+Y S P GY + P T+ + A Q+
 Sbjct: 117 IYMSDVTTRQPNWSTISSTN-PLWVAQYASMPNPTGYQ--SEFWTDGKGYGAWSSAAIHQY 173

Query: 175 TSKGRLPGYNGNLDLNVFYGDGNTWDLVYG 204
 +S G L ++GNLD+N+ Y + N W G
 Sbjct: 174 SSAGSLSNWNGNLDINLAYINANQNKSLAG 203

Query= sid|110164|lan|182ORF009 Phage 182 ORF|8765-9601|2
 (278 letters)

>gi|1429240|emb|CAA67659| (X99260) lower collar protein
 [Bacteriophage B103]
 Length = 293

Score = 180 bits (451), Expect = 1e-44

Identities = 115/296 (38%), Positives = 161/296 (53%), Gaps = 33/296 (11%)

Query: 3 LKRYIESPTYYPQLSRKERIEVGRKQLFDFDYFPYDETKRAEFETKFINHPLYREIGSE 62
 L YIE ++ Y+ LS E+IE GR +LFDY YP +DE+ R FET FI +FY+REIG E
 Sbjct: 8 LSTYIEMMSQYETGLSMAEKIEKGRPKLFDYFYPIFDESRYKVFETHPIRNFYMRIGFE 67

Query: 63 TMGSPKFNLDLYLNLNMPYWNKMFSLNLEEF-PIFDDMDYTIIDSKQLLNEDITNIKANR 121
 T G FKNL+ +L +NMPY+NK+P S L ++ P+ + T K+ DT NR
 Sbjct: 68 TEGLFKFNLETWLIINMPYFNKLFESELIKYDPLENTRLNNTTGNKKN-----DTERNDNR 122

Query: 122 D-----ESKNQTKQVDQTDNRNKNTRDTGTT-----DSFSRNTYTDTPQKDLRIASNG 169
 D + K+ TK D+T+ + D TT D+P+R +D P L + +N

338

Sbjct: 123 DTGSMKADGKSNKTSDKTNATGSSKEDGKTGTSVTDNDFNRKIDSQPDRLNLTTN- 181

Query: 170 DGTGVINYATNITEDLSKETTSSTGVETNNDKTNQNRSMAS-----EKETKNTD 219
 DG G + YA+ I E+ + ++TG TNN ++ + S S T N

Sbjct: 182 DGQGTLEYASAIEENNTNNKRNTTG--TNNVTSSAESESTGSGTSDVTITDNANTTTNDK 239

Query: 220 INKDQNTKDTITRYKGGKGNTRYADLLEKYRRSVLRIEKMIFREMKEGLFLLVY 275
 +N N +D I GK G YA L++ YR ++LRIEK IF EM + LF+LVY

Sbjct: 240 LNSQINNVEDYIESKIGSGTQSYASLVQDYRAALLRIEKRFDEMGE--LFMLVY 293

Query= sid|110165|lan|182ORF010 Phage 182 ORF|1310-2155|2
 (281 letters)

>gi|135604|sp|P06812|TERM_BPNF_DNA_TERMINAL_PROTEIN
 >gi|75815|pir|ERBPNP terminal protein - phage NF
 >gi|579177|emb|CAA68440| (Y00363) gene E product (AA
 1-267) [Bacteriophage NF]
 Length = 266

Score = 74.9 bits (181), Expect = 6e-13
 Identities = 73/275 (26%), Positives = 129/275 (46%), Gaps = 37/275 (13%)

Query: 3 VRISKNDRAKLEKIYGKSNKARKKYNRLRQK-GVE---ERQLPTVPTSKRLIDYVKSTN 58
 +RI+ ND+A K+ K+ KA K +R ++K G++ E +LP + + +

Sbjct: 7 IRTNNDKALYAKLV-KYTKA--KISRTKKKYGIDLSNEIPLPLESFQ----- 52

Query: 59 MSRSDFNMKMLDELVDPAQPYNENYIFEINKRNVAISRAQIKEAQIKTEQAQKAKEEHYKE 118
 +R +FNK + F N+NY F NK + S+A+I E T++AQ+ +E +E

Sbjct: 53 -TREEFNKWKQKQESFTNRAQNYQFVKNKYGIVASKAKINEIAKNTKEAQRIVDEQREE 111

Query: 119 L-----NKVEVKKPTENTIVTPTILTELGADLPFQAIPDFNIDAPTSPEGVQSYLEN 170
 + K + I++P+ +T G P DFN D S +++ E

Sbjct: 112 IEDKPFISGGKQGGTVGQRMQILSPSQVT--GISRP----SDFNFDVRSYARLRLTEEG 165

Query: 171 IG-KQDEQYFDERDQLYYDNFRQAMFTIFNSD--ADDIVRLDSMGLDLFMKTYVSNFLD 227
 + K Y+D R + NP + + FNSD +D+V L + D F + Y+ P +

Sbjct: 166 MAEKASPDYDRRMTQMHNQNFIEIVEKSFNSDWLDELVERLKKIPPDDFFELYLM-FDE 224

Query: 228 MNLDYIYDEAEVQKKEQVYSKIAKVIESETGGEV 262

++ +Y E E + E + +KI ++ G+V

Sbjct: 225 ISFEYFDEGEDVEASEAMLNKHISYLDYERGDV 259

Query= sid|110166|lan|182ORF011 Phage 182 ORF|9607-10158|1
 (183 letters)

>gi|1429241|emb|CAA67660| (X99260) pre-neck appendage protein
 [Bacteriophage B103]
 Length = 860

Score = 50.8 bits (119), Expect = 6e-06
 Identities = 29/105 (27%), Positives = 56/105 (52%), Gaps = 6/105 (5%)

Query: 8 KRFDGLPAVFKERFSKYPHTEYRYELLLDEEVSAIAYLNEVGALVNDMSGYLNIFYIEHF 67
 +RF+ L + + + +Y T + + L E+++ +I YLN++G L ND+ N +E

Sbjct: 7 RRFEKLGEMMVQVYERYLPTAFDESMTLLEKMMKIIEYLNQIGRLTNDVVEWKNVMEWI 66

Query: 68 V-EKLEETNDTLKKWLSGDTLENLINDTVFANYIKEIKRLQILV 111

++ +LE+ +TL+KM +G +L+ I E+K+ + V

Sbjct: 67 LNDGLEDYVKETLEKMYEKGKPADLV-----IQVIDELKQFGVSV 106

Query= sid|110169|lan|182ORF014 Phage 182 ORF|13716-14108|3
 (130 letters)

>gi|137936|sp|P11188|VG14_BPPH2_LYSIS_PROTEIN (LATE PROTEIN GP14)
 >gi|75860|pir|WMBP29 gene 14 protein - phage phi-29
 >gi|15678|emb|CAA28631| (X04962) gene 14 product (AA

339

1-393) [Bacteriophage phi-29] >gi|225369|prf||1301270J
 gene 14 [Bacteriophage phi-29]
 Length = 131

Score = 96.7 bits (237), Expect = 6e-20
 Identities = 53/131 (40%), Positives = 81/131 (61%), Gaps = 3/131 (2%)

Query: 1 MIEYITQWL-ADDNHLVYGLIWLVMAMIIDFVLGFTIAKFNKEIDFSSFKAKAGIIVKV 59
 MI ++ +L D+ L+Y L +LMV M++D VLG AK N I FSSFK K G++KV
 Sbjct: 3 MIAWMQHPLETDETKLIYWL-FLVMCMVVDTVLGVLPFAKLNPNIKFSSFKIKTGVLIKV 61

Query: 60 AEMVLVVYFIPVAVKFGAVGITMYITMLVGLILSEIYSILGHISDIDDDNNWTDYVKKFL 119
 +EM+L + IP AV F A G+ + T+ L +SEIYSI GH+ +DD +++ + ++ F
 Sbjct: 62 SEMILALLAIPFAVFPFA-GLPLLYTVYTALCVSEIYSIFGHLRLVDDKSDPLEILENPF 120

Query: 120 DGTLNRRKDDIK 130
 T + + K
 Sbjct: 121 KRTSGKNKEEK 131

Query= sid|110170|lan|182ORF015 Phage 182 ORF|854-1225|2
 (123 letters)

>gi|15670|emb|CAA24483| (V01155) reading frame 10 (may be gene 4)
 [Bacteriophage phi-29]
 Length = 124

Score = 69.9 bits (168), Expect = 6e-12
 Identities = 39/119 (32%), Positives = 64/119 (53%), Gaps = 3/119 (2%)

Query: 3 IVKSTFDTQTPEGMLQVFNATNGASIPLRNAI-GEVLELKDILVYSDEVSGFGGAEPSSQA 61
 IVK+TFDT+T EG +++FNA G +N G ++E I Y +G A+ +
 Sbjct: 6 IVKATFDTETLEQIKIFNAQTGGGQSFKNLPDGTIIEANAIAQYQVSDTYGDAK--EE 63

Query: 62 ELVAFFTEDGKTYAGVSAVATKSARKLIDMMTANPDIKPKISFVEGKSNNGGQKFNVLQV 120
 + F DG Y+ +S ++A +LID++T + K+ V+G S+ G P +LQ+
 Sbjct: 64 TVTITFAADGSLYSAISKTVAAASDLIDLVRHKLETFFKVVVQGTSSKGNVFFSLQL 122

Query= sid|110174|lan|182ORF019 Phage 182 ORF|4323-4613|3
 (96 letters)

>gi|1429235|emb|CAA67654| (X99260) head morphogenesis protein
 [Bacteriophage B103]
 Length = 101

Score = 60.9 bits (145), Expect = 1e-09
 Identities = 34/96 (35%), Positives = 53/96 (54%), Gaps = 5/96 (5%)

Query: 1 MEIKKHESILWGILESVDGEARSKIVEHLEALREDYGATTEALTSANSTLEKLKKNDA 60
 ME HE ILN + + + R+++ L+ LR DYG+ + S EKL+ +N
 Sbjct: 3 MERDSHEEILNKLNDPELHSETEL---LQQLRADYGSVLSEFSELTATEKLRAENS 59

Query: 61 LVISNSKLFREIRAIVEPAEN--NEPETDQNTLDDL 94
 L++SNSKLF+ I + E + E + IT++DL
 Sbjct: 60 LIVNSKLFQVVGITKEKEEIKQEELSETITIEDL 95

Query= sid|110180|lan|182ORF025 Phage 182 ORF|548-814|2
 (88 letters)

>gi|138099|sp|P06955|VG6 BPPZA EARLY PROTEIN GP6
 >gi|75841|pir|ERBP6Z gene 6 protein - phage PZA
 >gi|216047 (M11813) gene 6 product [Bacteriophage PZA]
 >gi|224746|prf||1112171K ORF 6 [Bacteriophage PZA]
 Length = 96

Score = 55.0 bits (130), Expect = 8e-08
 Identities = 28/79 (35%), Positives = 45/79 (56%)

340

Query: 4 KLMQRNVTSTKVEFSEVIVQDGAPTIVPCEPVVLTGKLSEKALSAIKRKNPDKNVVVN 63
K+MQR +T T V +++++ DG + G LS E+A +KRK + V V +
Sbjct: 3 KMMQREITKTTNVNVAKMVMVDGEVQVEQLPSETFVGNLSMEQAQWRMKRKYKGEPVQVVS 62

Query: 64 VSHETALYTMPVDKFIELD 82
V T +Y +PV+KF+E+A
Sbjct: 63 VEPNTEVYELPVEKFLEVA 81

Table 26

Secondary structure prediction for ORF 182ORF008

```

1  MMNGIDISSY QTGIDLSKVP CDFVNIKATG GTGYVNPDCD RAFQQALSLG KKIGVYHFAH
   CCCCCCCCCC CCCCCCCCCC CCEEEEEEEC CCCCCCCCCC HHHHHHHHHC CCCCEEEEEE
61  ERGLEGTPQQ EAQFFLDNIK GYIGKAVLIL DFEGSNQKDV NWAKAFLDYV YNKTGVKAWF
   CCCCCCCHH HHHHHHHHHC CCCCEEEEEE CCCCCCHHH HHHHHHHHHH HCCCCEEEE
121 YTYTANLNTT DFSSIAKGDY GLWVAEYGSN QPQGYSPAP PKTNNFPIVA CFQFTSKGRL
   EEECCCCCCC CCEEECCCCC CEEEEECCCC CCCCCCCCCC CCCCCCBEE EEEBCCCCC
181 PGYNGNLDLN VFYGDGNTWD LYVGKKQDQI VPPENKIFDA TSDEFIFTLT TGSTSVFYFD
   CCCCCCCEE EEECCCCCE EEECCCCCCC CCCCCCCCCC CCCEEEEEEC CCCCEEEEC
241 GETIFELSDP TQLDHIRGTY NHVHGKEIPS MWTPEQFDI YLKMYEKKPV YK
   CCEEECCCC CCHHHHCCEE CCCCCCEEC CCCCCCHHH HHHHHCCCC EC

```

Secondary structure prediction for ORF 182ORF014

```

1  MIEYITQWLA DDNHLVYGLI IWLVMAMIID FVLGFTIAKF NKEIDFSSFK AKAGIIVKVA
   CCCCEEECCC CCCCHHHHHH HHHHHHHHHH HHHHHHHHHC CCCCCHHHH HHHCEEEERE
61  EMVLVVYFIP VAVKFGAVGI TMYITMLVGL ILSEIYSILG HISDIDDDNN WTDYVKKFLD
   EEEEEEEEC CEEECCEEE EEEEEEEEEE EEEEEEEEC CCCCCCCCC CEEEEEEECC
121 GTLNRKDDIK
   CCCCCCEEC

```

Table 27

Enterococcus accession numbers 242/242

gi 2895751 gb AF044978.1 AF044978 [2895751]	gi 4098267 gb U76614.1 BLU76614 [4098267]
gi 4803755 dbj AB026843.1 AB026843 [4803755]	gi 47019 emb Y00116.1 SFAMB1 [47019]
gi 4769001 gb AF140549.1 AF140549 [4769001]	gi 4158179 emb AL035206.1 SC9B5 [4158179]
gi 4760901 gb AF099088.1 AF099088 [4760901]	gi 4165458 emb X79343.1 EF16SSPA [4165458]
gi 4704705 gb AF121254.1 AF121254 [4704705]	gi 4165457 emb X79342.1 EFTRNALA [4165457]
gi 3342117 gb AF076604.1 AF076604 [3342117]	gi 4165456 emb X79341.1 EF23SRNA [4165456]
gi 4688824 emb AJ132470.1 ESP132470 [4688824]	gi 4150978 emb Y14027.1 EFY14027 [4150978]
gi 4732085 gb AF125553.1 AF125553 [4732085]	gi 4127803 emb AJ223161.1 EFAJ3161 [4127803]
gi 4732082 gb AF125552.1 AF125552 [4732082]	gi 2956685 emb Y16413.1 EFENTIJO [2956685]
gi 4732079 gb AF125551.1 AF125551 [4732079]	gi 2665346 emb Y13922.1 EHY13922 [2665346]
gi 4732076 gb AF125550.1 AF125550 [4732076]	gi 4324675 gb AF109375.1 AF109375 [4324675]
gi 4732073 gb AF125548.1 AF125548 [4732073]	gi 4234627 gb AF061013.1 AF061013 [4234627]
gi 4732070 gb AF125547.1 AF125547 [4732070]	gi 4234626 gb AF061012.1 AF061012 [4234626]
gi 4732067 gb AF125546.1 AF125546 [4732067]	gi 4234625 gb AF061011.1 AF061011 [4234625]
gi 4732064 gb AF125545.1 AF125545 [4732064]	gi 4234624 gb AF061010.1 AF061010 [4234624]
gi 4732061 gb AF125544.1 AF125544 [4732061]	gi 4234623 gb AF061009.1 AF061009 [4234623]
gi 4704653 gb AF114715.1 AF114715 [4704653]	gi 4234622 gb AF061008.1 AF061008 [4234622]
gi 4704564 gb AF102550.1 AF102550 [4704564]	gi 4234621 gb AF061007.1 AF061007 [4234621]
gi 4688827 emb AJ238249.1 EFA238249 [4688827]	gi 4234620 gb AF061006.1 AF061006 [4234620]
gi 4680606 gb AF125198.1 AF125198 [4680606]	gi 4234619 gb AF061005.1 AF061005 [4234619]
gi 4633279 gb AF117609.1 AF117609 [4633279]	gi 4234618 gb AF061004.1 AF061004 [4234618]
gi 4633124 gb AF110130.1 AF110130 [4633124]	gi 4234617 gb AF061003.1 AF061003 [4234617]
gi 4590399 gb AF124258.1 AF124258 [4590399]	gi 4234616 gb AF061002.1 AF061002 [4234616]
gi 4590336 gb AF108380.1 AF108380 [4590336]	gi 4234615 gb AF061001.1 AF061001 [4234615]
gi 4590335 gb AF108379.1 AF108379 [4590335]	gi 4234614 gb AF061000.1 AF061000 [4234614]
gi 4019167 gb U21300.1 CXU21300 [4019167]	gi 3138990 gb AF060241.1 AF060241 [3138990]
gi 4545122 gb AF077816.1 AF077816 [4545122]	gi 3138986 gb AF060240.1 AF060240 [3138986]
gi 4433610 gb AF106614.1 AF106614 [4433610]	gi 4204535 gb AF094803.1 AF094803 [4204535]
gi 4468838 emb AJ132039.1 EFA132039 [4468838]	gi 4204534 gb AF094802.1 AF094802 [4204534]
gi 4468121 emb AJ132958.1 BPH132958 [4468121]	gi 4204533 gb AF094801.1 AF094801 [4204533]
gi 4456104 emb Y17302.1 EHI17302 [4456104]	gi 4204532 gb AF094800.1 AF094800 [4204532]
gi 4433611 gb AF106615.1 AF106615 [4433611]	gi 4204531 gb AF094799.1 AF094799 [4204531]
gi 4433607 gb AF106611.1 AF106611 [4433607]	gi 4204530 gb AF094798.1 AF094798 [4204530]
	gi 4204529 gb AF094797.1 AF094797 [4204529]
	gi 4204528 gb AF094796.1 AF094796 [4204528]
	gi 4204527 gb AF094795.1 AF094795 [4204527]

gi|4204526|gb|AF094794.1|AF094794 [4204526]
 gi|4204525|gb|AF094793.1|AF094793 [4204525]
 gi|4204524|gb|AF094792.1|AF094792 [4204524]
 gi|4204523|gb|AF094791.1|AF094791 [4204523]
 gi|4204522|gb|AF094790.1|AF094790 [4204522]
 gi|4204521|gb|AF094789.1|AF094789 [4204521]
 gi|4204520|gb|AF094788.1|AF094788 [4204520]
 gi|4204519|gb|AF094787.1|AF094787 [4204519]
 gi|4204518|gb|AF094786.1|AF094786 [4204518]
 gi|4204517|gb|AF094785.1|AF094785 [4204517]
 gi|4204516|gb|AF094784.1|AF094784 [4204516]
 gi|4204515|gb|AF094783.1|AF094783 [4204515]
 gi|4204514|gb|AF094782.1|AF094782 [4204514]
 gi|4204513|gb|AF094781.1|AF094781 [4204513]
 gi|4204512|gb|AF094780.1|AF094780 [4204512]
 gi|3873186|gb|AF034779.1|AF034779 [3873186]
 gi|4151367|gb|AF093508.1|AF093508 [4151367]
 gi|2828136|gb|AF039903.1|AF039903 [2828136]
 gi|2828135|gb|AF039902.1|AF039902 [2828135]
 gi|2828134|gb|AF039901.1|AF039901 [2828134]
 gi|2828133|gb|AF039900.1|AF039900 [2828133]
 gi|2828132|gb|AF039899.1|AF039899 [2828132]
 gi|2828131|gb|AF039898.1|AF039898 [2828131]
 gi|4103866|gb|AF028812.1|AF028812 [4103866]
 gi|4103864|gb|AF028811.1|AF028811 [4103864]
 gi|2605925|gb|AF029727.1|AF029727 [2605925]
 gi|1402750|gb|U60038.1|EFU60038 [1402750]
 gi|1835780|gb|U86375.1|EFU86375 [1835780]
 gi|3831555|gb|AF047608.1|AF047608 [3831555]
 gi|3790617|gb|AF097414.1|AF097414 [3790617]
 gi|3767587|dbj|AB005036.1|AB005036 [3767587]
 gi|3757810|gb|AF042288.1|AF042288 [3757810]
 gi|3747039|gb|AF093509.1|AF093509 [3747039]
 gi|3660559|dbj|AB017811.1|AB017811 [3660559]
 gi|1147743|gb|U42211.1|EHU42211 [1147743]
 gi|3676412|gb|AF051917.1|AF051917 [3676412]
 gi|3676164|emb|AJ011113.1|EFA011113 [3676164]
 gi|2612869|gb|AF005726.1|AF005726 [2612869]
 gi|2353762|gb|AF016233.1|AF016233 [2353762]
 gi|2149899|gb|U94707.1|EFU94707 [2149899]
 gi|2149149|gb|U82366.1|LSU82366 [2149149]
 gi|1469463|gb|U49512.1|EFU49512 [1469463]
 gi|1244503|gb|U35366.1|EFU35366 [1244503]
 gi|833854|gb|U26268.1|EFU26268 [833854]
 gi|841200|gb|U18931.1|CPU18931 [841200]
 gi|460079|gb|U00457.1|U00457 [460079]
 gi|460077|gb|U00456.1|U00456 [460077]
 gi|535661|gb|L34675.1|INSTRANSPO [535661]
 gi|3023041|gb|AF007787.1|AF007787 [3023041]
 gi|431124|gb|L15633.1|TRN916ENT [431124]
 gi|388106|gb|L23802.1|ENEEBSA [388106]
 gi|3608387|gb|AF071085.1|AF071085 [3608387]
 gi|3551851|gb|AF076027.1|AF076027 [3551851]
 gi|3551773|gb|U94770.1|SPU94770 [3551773]
 gi|3551743|gb|U57498.1|ECU57498 [3551743]
 gi|3243178|gb|AF063010.1|AF063010 [3243178]
 gi|3136316|gb|AF063900.1|AF063900 [3136316]
 gi|3540256|gb|AF052459.1|AF052459 [3540256]
 gi|755215|gb|U17696.1|LLU17696 [755215]
 gi|3421437|gb|AF082295.1|AF082295 [3421437]
 gi|3421436|gb|AF082294.1|AF082294 [3421436]
 gi|3421435|gb|AF082293.1|AF082293 [3421435]
 gi|3421434|gb|AF082292.1|AF082292 [3421434]
 gi|3341430|emb|Y17797.1|EFY17797 [3341430]
 gi|3319647|emb|X69092.1|EHPBP3RA [3319647]
 gi|3292886|emb|AJ007584.1|EFA7584 [3292886]
 gi|3261536|emb|AL021958.1|MTV041 [3261536]
 gi|3250708|emb|Z95150.1|MTCY164 [3250708]
 gi|3249688|gb|AF070678.1|AF070678 [3249688]
 gi|3249687|gb|AF070677.1|AF070677 [3249687]
 gi|3249686|gb|AF070676.1|AF070676 [3249686]
 gi|3219158|dbj|AB015233.1|AB015233 [3219158]
 gi|2765275|emb|Y12924.1|SPY12924 [2765275]
 gi|3183687|emb|Y11621.1|EA16SRRN [3183687]
 gi|2765274|emb|Y12923.1|EFY12923 [2765274]
 gi|2765273|emb|Y12922.1|ESY12922 [2765273]
 gi|2765272|emb|Y12921.1|ESY12921 [2765272]
 gi|2765271|emb|Y12920.1|EDY12920 [2765271]
 gi|2765270|emb|Y12919.1|ESY12919 [2765270]

gi|2765269|emb|Y12918.1|ECY12918 [2765269]
 gi|2765268|emb|Y12917.1|ECY12917 [2765268]
 gi|2765267|emb|Y12916.1|EPY12916 [2765267]
 gi|2765266|emb|Y12915.1|ESY12915 [2765266]
 gi|2765265|emb|Y12914.1|ERY12914 [2765265]
 gi|2765264|emb|Y12913.1|EMY12913 [2765264]
 gi|2765263|emb|Y12912.1|EHY12912 [2765263]
 gi|2765262|emb|Y12911.1|EMY12911 [2765262]
 gi|2765261|emb|Y12910.1|EGY12910 [2765261]
 gi|2765260|emb|Y12909.1|EDY12909 [2765260]
 gi|2765259|emb|Y12908.1|ECY12908 [2765259]
 gi|2765258|emb|Y12907.1|EAY12907 [2765258]
 gi|2765257|emb|Y12906.1|EFY12906 [2765257]
 gi|2765256|emb|Y12905.1|EFY12905 [2765256]
 gi|2894541|emb|AJ223332.1|EFAJ3332 [2894541]
 gi|2894539|emb|AJ223331.1|EFAJ3331 [2894539]
 gi|3108058|gb|AF060881.1|AF060881 [3108058]
 gi|3087776|emb|AJ223633.1|EFAJ3633 [3087776]
 gi|3080754|gb|AF016483.1|AF016483 [3080754]
 gi|2197119|gb|AF003921.1|AF003921 [2197119]
 gi|2982722|dbj|AB012213.1|AB012213 [2982722]
 gi|2982721|dbj|AB012212.1|AB012212 [2982721]
 gi|2058780|gb|B07890.1|B07890 [2058780]
 gi|2058779|gb|B07889.1|B07889 [2058779]
 gi|2058778|gb|B07888.1|B07888 [2058778]
 gi|2058777|gb|B07887.1|B07887 [2058777]
 gi|2058776|gb|B07886.1|B07886 [2058776]
 gi|2058775|gb|B07885.1|B07885 [2058775]
 gi|2058774|gb|B07884.1|B07884 [2058774]
 gi|2058773|gb|B07873.1|B07873 [2058773]
 gi|2058772|gb|B07872.1|B07872 [2058772]
 gi|2058771|gb|B07871.1|B07871 [2058771]
 gi|2058770|gb|B07870.1|B07870 [2058770]
 gi|2058769|gb|B07869.1|B07869 [2058769]
 gi|2058768|gb|B07868.1|B07868 [2058768]
 gi|2058767|gb|B07867.1|B07867 [2058767]
 gi|2058766|gb|B07866.1|B07866 [2058766]
 gi|2058765|gb|B07865.1|B07865 [2058765]
 gi|2058764|gb|B07864.1|B07864 [2058764]
 gi|2058763|gb|B07883.1|B07883 [2058763]
 gi|2058762|gb|B07882.1|B07882 [2058762]
 gi|2058761|gb|B07881.1|B07881 [2058761]
 gi|2058760|gb|B07880.1|B07880 [2058760]
 gi|2058759|gb|B07879.1|B07879 [2058759]
 gi|2058758|gb|B07878.1|B07878 [2058758]
 gi|2058757|gb|B07877.1|B07877 [2058757]
 gi|2058756|gb|B07876.1|B07876 [2058756]
 gi|2058755|gb|B07875.1|B07875 [2058755]
 gi|2058754|gb|B07874.1|B07874 [2058754]
 gi|2058753|gb|B07863.1|B07863 [2058753]
 gi|2058752|gb|B07862.1|B07862 [2058752]
 gi|2058751|gb|B07861.1|B07861 [2058751]
 gi|2058750|gb|B07860.1|B07860 [2058750]
 gi|2058749|gb|B07859.1|B07859 [2058749]
 gi|2058748|gb|B07858.1|B07858 [2058748]
 gi|2058747|gb|B07857.1|B07857 [2058747]
 gi|2058746|gb|B07856.1|B07856 [2058746]
 gi|2058745|gb|B07855.1|B07855 [2058745]
 gi|2058744|gb|B07854.1|B07854 [2058744]
 gi|2058743|gb|B07853.1|B07853 [2058743]
 gi|2058742|gb|B07852.1|B07852 [2058742]
 gi|2058741|gb|B07851.1|B07851 [2058741]
 gi|2058740|gb|B07850.1|B07850 [2058740]
 gi|2947527|gb|T25933.1|T25933 [2947527]
 gi|2924302|emb|X81655.1|EHERMAM [2924302]
 gi|2664256|emb|Y12234.1|EFAS48C [2664256]
 gi|2879906|dbj|D85752.1|D85752 [2879906]
 gi|2746216|gb|AF028836.1|AF028836 [2746216]
 gi|2745825|gb|AF039139.1|AF039139 [2745825]
 gi|2696019|dbj|AB007844.1|AB007844 [2696019]
 gi|48999|emb|X62280.1|EHPBP5G [48999]
 gi|2654477|gb|U89914.1|BFU89914 [2654477]
 gi|43347|emb|X68646.1|EHPSRAA [43347]
 gi|2613034|gb|AH005624.1|SEG_EDDH4RR [2613034]
 gi|2613033|gb|AF029775.1|EDDH4RR2 [2613033]
 gi|2613032|gb|AF029774.1|EDDH4RR1 [2613032]
 gi|2613031|gb|AH005623.1|SEG_EDDH1RR [2613031]
 gi|2613030|gb|AF029773.1|EDDH1RR2 [2613030]

gi|2613029|gb|AF029772.1|EDDHIRR1 [2613029]
 gi|2613028|gb|AH005622.1|SEG_EDH19RR
 [2613028]
 gi|2613027|gb|AF029771.1|EDH19RR2 [2613027]
 gi|2613026|gb|AF029770.1|EDH19RR1 [2613026]
 gi|2613025|gb|AH005621.1|SEG_EDISRR
 [2613025]
 gi|2613024|gb|AF029769.1|EDISRR2 [2613024]
 gi|2613023|gb|AF029768.1|EDISRR1 [2613023]
 gi|1881226|dbj|AB001488.1|AB001488 [1881226]
 gi|2547160|gb|AF023104.1|AF023104 [2547160]
 gi|2547159|gb|AF023103.1|AF023103 [2547159]
 gi|2547158|gb|AF023102.1|AF023102 [2547158]
 gi|2547157|gb|AF023101.1|AF023101 [2547157]
 gi|2415383|gb|AF015775.1|AF015775 [2415383]
 gi|2388636|gb|U94356.1|EFU94356 [2388636]
 gi|2388634|gb|U94355.1|ECU94355 [2388634]
 gi|2340825|dbj|D26045.1|D26045 [2340825]
 gi|2226147|emb|Y14080.1|BSY14080 [2226147]
 gi|2327026|gb|U87997.1|EFU87997 [2327026]
 gi|2318058|gb|AF012532.1|AF012532 [2318058]
 gi|1848175|emb|X87189.1|EM23S5SSP [1848175]
 gi|1848174|emb|X87187.1|EM16S23SS [1848174]
 gi|1848173|emb|X87188.1|EM16S23SP [1848173]
 gi|1848172|emb|X87185.1|EH23S5SSP [1848172]
 gi|1848171|emb|X87184.1|EH16S23SS [1848171]
 gi|1848170|emb|X87181.1|EF23S5SSP [1848170]
 gi|1848169|emb|X87183.1|EF23S5SPA [1848169]
 gi|1848168|emb|X87191.1|EF23S5SAC [1848168]
 gi|1848167|emb|X87180.1|EF16S23SS [1848167]
 gi|1848166|emb|X87182.1|EF16S23SP [1848166]
 gi|1848165|emb|X87190.1|EF16S23SC [1848165]
 gi|1848164|emb|X87186.1|EF16S23SA [1848164]
 gi|1848156|emb|X87179.1|ED23S5SSP [1848156]
 gi|1848155|emb|X87178.1|ED16S23SS [1848155]
 gi|1848154|emb|X87177.1|ED16S23SA [1848154]
 gi|2274942|emb|AJ000346.1|EHNAPBC [2274942]
 gi|2274939|emb|AJ000042.1|EFGLS24B
 [2274939]
 gi|414575|gb|L12710.1|ENEAAC [414575]
 gi|2245603|gb|AF006008.1|AF006008 [2245603]
 gi|2231992|gb|U94530.1|EFU94530 [2231992]
 gi|2231990|gb|U94529.1|EFU94529 [2231990]
 gi|2231988|gb|U94528.1|EFU94528 [2231988]
 gi|2231986|gb|U94527.1|EFU94527 [2231986]
 gi|2231984|gb|U94526.1|EFU94526 [2231984]
 gi|2231982|gb|U94525.1|ECU94525 [2231982]
 gi|2231980|gb|U94524.1|ECU94524 [2231980]
 gi|2231978|gb|U94523.1|ECU94523 [2231978]
 gi|2231976|gb|U94522.1|ECU94522 [2231976]
 gi|2231974|gb|U94521.1|ECU94521 [2231974]
 gi|2196685|gb|U25090.1|EFU25090 [2196685]
 gi|2197120|gb|AF003922.1|AF003922 [2197120]
 gi|2196683|gb|U25095.1|EFU25095 [2196683]
 gi|2196681|gb|U25094.1|EFU25094 [2196681]
 gi|2196679|gb|U25093.1|EFU25093 [2196679]
 gi|2196677|gb|U25092.1|EFU25092 [2196677]
 gi|2196675|gb|U25091.1|EFU25091 [2196675]
 gi|2196673|gb|U24682.1|EFU24682 [2196673]
 gi|532533|gb|U09422.1|EFU09422 [532533]
 gi|487271|dbj|D17462.1|ENENTP [487271]
 gi|468459|dbj|D28859.1|ENEPPD1 [468459]
 gi|440135|dbj|D16334.1|ENEATPK [440135]
 gi|391680|dbj|D13816.1|ENENAABS [391680]
 gi|1402524|dbj|D78257.1|D78257 [1402524]
 gi|709995|dbj|D30808.1|BACYCB20 [709995]
 gi|2109265|gb|U91527.1|EFU91527 [2109265]
 gi|1041112|dbj|D78016.1|ENEPPD1A [1041112]
 gi|1339880|dbj|D85392.1|ENERPA [1339880]
 gi|1339878|dbj|D85393.1|ENEGE1E [1339878]
 gi|662918|emb|Z46807.1|EHCOPAYZ [662918]
 gi|769796|emb|X86176.1|EFRPODDNE [769796]
 gi|1854638|gb|U51479.1|EGU51479 [1854638]
 gi|1857221|gb|U72706.1|EFU72706 [1857221]
 gi|1857219|gb|U72704.1|EFU72704 [1857219]
 gi|1857217|gb|U72705.1|ECU72705 [1857217]
 gi|1272655|emb|X96978.1|EFPPD1GN\$ [1272655]
 gi|1272652|emb|X96976.1|EFPLSEF1G [1272652]
 gi|1279406|emb|X96977.1|EFPAD1ORF
 [1279406]
 gi|1070149|emb|X93211.1|EFTNFO1 [1070149]

gi|1065723|emb|X92947.1|EFTETMGN [1065723]
 gi|1019639|gb|L38972.1|PH4COINJN [1019639]
 gi|1151151|gb|U43087.1|EFU43087 [1151151]
 gi|1098507|gb|U17283.1|BMU17283 [1098507]
 gi|1498072|gb|U64887.1|EFU64887 [1498072]
 gi|1498071|gb|U64886.1|EFU64886 [1498071]
 gi|1469783|gb|U58049.1|EHU58049 [1469783]
 gi|1763666|gb|U81452.1|EFU81452 [1763666]
 gi|624694|gb|L38973.1|PH4SEQ [624694]
 gi|1730458|emb|Z83305.1|EFVANRES [1730458]
 gi|1419498|emb|X84796.1|ECPFW4 [1419498]
 gi|1419497|emb|X84795.1|ECPFW3 [1419497]
 gi|1419496|emb|X84794.1|ECPFW1 [1419496]
 gi|254400|gb|S43266.1|S43266 [254400]
 gi|239025|gb|S66277.1|S66277 [239025]
 gi|1054931|gb|U38590.1|EFU38590 [1054931]
 gi|1244573|gb|U39788.1|EHU39788 [1244573]
 gi|1244571|gb|U39789.1|EGU39789 [1244571]
 gi|1244569|gb|U39790.1|EFU39790 [1244569]
 gi|1255020|gb|U39777.1|ESU39777 [1255020]
 gi|1255018|gb|U39775.1|EPU39775 [1255018]
 gi|1255016|gb|U39778.1|EDU39778 [1255016]
 gi|1255014|gb|U39776.1|ECU39776 [1255014]
 gi|1255012|gb|U39774.1|EAU39774 [1255012]
 gi|1619922|gb|U69267.1|IVU69267 [1619922]
 gi|790436|emb|X84861.1|EFEFMPBP5 [790436]
 gi|790434|emb|X84858.1|EFD63RPSR [790434]
 gi|790432|emb|X84862.1|EF721PBP5 [790432]
 gi|790430|emb|X84860.1|EF63RPBP5 [790430]
 gi|790428|emb|X84859.1|EF366PBP5 [790428]
 gi|1572800|gb|U70854.1|CELF38A5 [1572800]
 gi|1041816|gb|U17153.1|EFU17153 [1041816]
 gi|1086523|gb|U39859.1|EFU39859 [1086523]
 gi|403564|gb|U01917.1|EFU01917 [403564]
 gi|1515474|gb|U66286.1|EFU66286 [1515474]
 gi|1513068|gb|U15554.1|LMU15554 [1513068]
 gi|1296520|emb|X94181.1|EFENTAORF
 [1296520]
 gi|1488069|gb|U63997.1|EFU63997 [1488069]
 gi|1209525|gb|U35369.1|EFU35369 [1209525]

gi|1469341|gb|U30931.1|ESU30931 [1469341]
 gi|488331|gb|M77276.1|SYNGIP2122 [488331]
 gi|1046177|gb|U39733.1| [1046177]
 gi|1236613|gb|U49939.1|CVU49939 [1236613]
 gi|47491|emb|X55766.1|SS16SR5G [47491]
 gi|47490|emb|X55767.1|SS16SR3G [47490]
 gi|47061|emb|X56353.1|SFTET916 [47061]
 gi|49022|emb|X62755.1|SFNPRG [49022]
 gi|47047|emb|X17214.1|SFPASA1 [47047]
 gi|47044|emb|X68847.1|SFNOXAA [47044]
 gi|47033|emb|V01547.1|SFKANR [47033]
 gi|47018|emb|X02027.1|SF5SRNA [47018]
 gi|511044|emb|X75752.1|MP16SRNA0 [511044]
 gi|511043|emb|X75751.1|MP16SR243 [511043]
 gi|886481|emb|X82819.1|ESPLPAM [886481]
 gi|517387|emb|X76177.1|ES16SRR [517387]
 gi|472916|emb|X76913.1|EHNTPOP [472916]
 gi|43351|emb|X55133.1|ES16SRRN [43351]
 gi|1143442|emb|X92687.1|EFPBP5G [1143442]
 gi|963032|emb|Z50854.1|EHARPQTOU [963032]
 gi|886479|emb|X84818.1|EHDNAPSR [886479]
 gi|551437|emb|X81654.1|EHIS1216 [551437]
 gi|467805|emb|X78425.1|EFPBP5 [467805]
 gi|296721|emb|X55961.1|EFPD78 [296721]
 gi|287946|emb|Z19137.1|EFPTSHGN [287946]
 gi|49042|emb|X63285.1|EHNKA [49042]
 gi|49019|emb|X62658.1|EFSEA1 [49019]
 gi|43337|emb|Z12296.1|EFSPREG [43337]
 gi|43335|emb|X56895.1|EFPVANAG [43335]
 gi|43333|emb|X16421.1|EFPF54 [43333]
 gi|43331|emb|X62657.1|EFORF3 [43331]
 gi|1065721|emb|X92945.1|EFCAT501 [1065721]
 gi|806551|emb|Z49243.1|EF4110SOD [806551]
 gi|806549|emb|Z49244.1|EF410SOD [806549]
 gi|505530|emb|X79542.1|EFAS48 [505530]
 gi|43323|emb|X62656.1|EFASPT [43323]
 gi|40840|emb|X56422.1|EC16SRNAG [40840]
 gi|48189|emb|X04388.1|TN1545TR [48189]
 gi|928814|gb|L40841.1|ENETRANSPO [928814]
 gi|141856|gb|L01794.1|AD1REPABC [141856]

gi|149125|gb|M90647.1|IP8VANY [149125]
gi|141862|gb|M87836.1|AD1TRAE1 [141862]
gi|141860|gb|M84374.1|AD1TRAA [141860]
gi|141853|gb|M62888.1|AD1PAD1 [141853]
gi|1101637|dbj|D31674.1|EVM16RNA7 [1101637]
gi|1101636|dbj|D31675.1|ENE16RNA8 [1101636]
gi|497792|dbj|D31676.1|ENC16RNA9 [497792]
gi|1022729|gb|U36195.1|EFU36195 [1022729]
gi|488338|gb|M77279.1|SYNGIP3124 [488338]
gi|488335|gb|M77278.1|SYNGIP2563 [488335]
gi|488333|gb|M77277.1|SYNGIP2124 [488333]
gi|488329|gb|M77275.1|SYNGIP2121 [488329]
gi|388267|gb|L19532.1|AD1TRAC [388267]
gi|493016|gb|U03756.1|EFU03756 [493016]
gi|453536|gb|L28754.1|INSTRAN [453536]
gi|153658|gb|M58002.1|STRHYDROLA [153658]
gi|475427|gb|U00681.1|EFU00681 [475427]
gi|818704|gb|U24692.1|EFU24692 [818704]
gi|155036|gb|M97297.1|TRNVAN [155036]
gi|150552|gb|M64978.1|PCFPRGAB [150552]
gi|786274|gb|U22541.1|EHU22541 [786274]
gi|786273|gb|U22540.1|EHU22540 [786273]
gi|559858|gb|L37110.1|AD1CLYL [559858]
gi|643614|gb|U16659.1|ECU16659 [643614]
gi|643612|gb|U16658.1|ECU16658 [643612]
gi|290641|gb|L13292.1|ENECOPPUMP [290641]
gi|624701|gb|L29639.1|ENEVANCRF [624701]
gi|624699|gb|L29638.1|ENEVANC [624699]
gi|624692|gb|L29641.1|ENEDDLA [624692]
gi|624690|gb|L29640.1|ENEDDL [624690]
gi|493094|gb|L32813.1|ENERRD [493094]

gi|153852|gb|AH000939.1|SEG_STRTN916 [153852]
gi|153851|gb|M22645.1|STRTN9162 [153851]
gi|153850|gb|M20864.1|STRTN9161 [153850]
gi|153660|gb|M36878.1|STRIF2BA [153660]
gi|153585|gb|M13771.1|STRBRP [153585]
gi|153575|gb|M64265.1|STRATPEFHA [153575]
gi|153565|gb|M90060.1|STRATPASEA [153565]
gi|152969|gb|M92376.1|STABLAIA [152969]
gi|309660|gb|L14285.1|PCFPRGWZY [309660]
gi|433714|gb|L12033.1|ENESATA [433714]
gi|290645|gb|L15304.1|ENEVANB2A [290645]
gi|148331|gb|M84146.1|ENEVANR [148331]
gi|148329|gb|M64304.1|ENEVANH [148329]
gi|148326|gb|M68910.1|ENEVANCRES [148326]
gi|148324|gb|M75132.1|ENEVANC [148324]
gi|148323|gb|L06138.1|ENEVANB [148323]
gi|148321|gb|M85225.1|ENETETM [148321]
gi|148320|gb|L00925.1|ENERTRNA [148320]
gi|148319|gb|L00924.1|ENERRNA [148319]
gi|148317|gb|M81466.1|ENERECA [148317]
gi|148315|gb|M81961.1|ENENAPA [148315]
gi|148312|gb|M38386.1|ENEMSPDPS [148312]
gi|148310|gb|M37185.1|ENEGELE [148310]
gi|148307|gb|L07892.1|ENEBLACREG [148307]
gi|148305|gb|M60253.1|ENEBELAA [148305]
gi|148303|gb|M77639.1|ENEB14NAM [148303]
gi|290644|gb|L16515.1|ENERGTG [290644]
gi|154954|gb|M37184.1|TRN916 [154954]
gi|148301|gb|M69221.1|ENEAAD9A [148301]
gi|148308|gb|M38052.1|ENECYLB [148308]

Table 28

Phage Dp1 complete genome sequence. 56506 nucleotides.

```

1      ataataaaaa tatgaagcag atattggggt aartatttgc taacaaaatg caccgaattt gtgtataata
71     taagtgaagc agttttgtaa acctgacatc ctgctaataa aaaataaagg aggcctcgaac atgagtcaaa
141    acactacacg cactgacgct gaattgacag cgcgttactct tttaggaaac caagacacca aatcacgatta
211    tgactataat ccagacgtcc ttgaaacttt ccctaacaaa catcctgaaa ataattacct agtaacattt
281    gacggatagc aattcacttc cctttgcctt aaaacaggac agcctgactt cgcgaatgtt ttcattagtt
351    acattccaaa cgaaaagatg gttgaatcta aatcattgaa attgtactta ttcagtttcc gtaaccacgg
421    tgacttccac gaagattgca tgaacattat ttgaaargac ttgtatgaat tgatggaaac taagtacatt
491    gaagtcacgg gcctattcac tcttcgtggt ggaatttcaa tttaccattt cgtcaacaaa gtgaatcctc
561    aatttgcaac tcttgaacct gaacagcttc aacttcaacg caaattgaac ttccttggaa atgttcaagg
631    tcttggacga gctattcgat aggaggtcgg aatgaaatca gtatgtttat tatccggcgg agtcgactca
701    gccacttgtt tagcaattga agttgacaag tgggggttcta aaaaatgttca tgctatagca ttcattacg
771    gacaaaagca tgaagcagaa ctgaaaatg ctgctaattg tgcattgttc tacggagtc agttcaccat
841    tcttgaaatt gactcgaaaa tctactcaag ctctagctct tecttattac aaggaaaaag cgaaatttca
911    catggaaaat cttacgctga aatccttagca gagaagggaag tagttgacac ctatgttcca tttagaattt
981    gactaatgct ttcacagggt cgggcttatg cttattcggg tggagcttct tacgtcgtat atggtgctca
1051   cgcagacgat cgggctggag gtgcttaccg tgattgcact cctgagttct ataattcaat gtcataatgca
1121   atggaatagc gaactggagg caaggtaacc cttgtcgtct ctctacttac tctaaccaag gcgcaagtgc
1191   ttaaatgggg aattgattta gatgttctct atttcttgac tctgtcatgt tatgaaagtg acgtgaaag
1261   ttgtggaact tgcgcaactt gtatcgaccg caaaaaggca ttcgaagaaa atggaatgac tgacctattt
1331   cattataagg agaattgata tgagagtttc taaaacctta acattcgacg cagctcatca actagttgga
1401   cattttggaa aatgcgcaaa tttgcacggg cataacttaca aagtcgaaat ttcattagca ggcggaactt
1471   atgaccacgg ttcgagtcac gggatgggtt ttgactttta tcaactcgaag aaaaatcgag gtacattcat
1541   tgacagactt gaccacgctg tcttcttcca agggaaatgaa ccaatcgctt tagcaaatgc agttgacacc
1611   aagcgagttc tatttggatt tagaactacg cgtgagaata tctcaagatt ccttacctgg actctacgg
1681   agcttatgtg gaagcatgct cgtatcgact ctatcaaaat atgggaaact cctacaggtt gcgcagaatg
1751   tacttactac gagattttca cagaagcaga gattgaaatg tcaagaacg taacctttat cgacaaagac
1821   gaaaagatta ctgtccgcga aatttttagag caggagcagg ataattggtt atcaatacaa tcagcctgaa
1891   agaggcaaga ttcgaatcaa tgttcgcgac cctgagaaaa tgcttatcat ggaaatttct gggtctacaa
1961   ttcgaagtga aggaatgggt ataggtcaca agactatttt cattcgaact ggtggatgag actatcattg
2031   caactgggtg gactcagcct ttacctggaa cgggtactact gagccggaat atatcacagg caaagaagct
2101   gctagtcgaa tcttgaactt agctttcaat gataaagggt aacagatttg taaccacgtg acattgactg
2171   gaggaatccc tgcccttaate aacgagccta tggctaaagt gatttcgatt ctaaaagaac atggattcaa
2241   gtttgggttc gaaactcaag gaactcgatt ccaagaatgg ttcagaagag taagcgatat cactattagt
2311   ccttaaacgc cttcaagtgg aatgagaact aatatgaaaa ttcttgaagc tatgttagat agaattgaatg
2381   atgaaaacct tgacttggtc tttaaaatcg ttatcttcca cgaaaatgac ctatgttatg cgcgtgatat
2451   gtttaaaact ttcgaaggca agttacgtcc agtgaactac ctttcagttg ggaatgcaaa cgcatacgaa
2521   gaaggaaaaa tcagtgatag gtttcttgaa aagttgggat ggctttggga taaagtgtat gaagacccag
2591   ctttcaacaa tgttcgacct ttacgcgaac ttcatacact tgtttatgat aataaaagag gattatataa
2661   tgaaaattga gcactctgat aaaatcggta acgtattagg gagagagaac ggaatgggctt cctctaagcc
2731   ggaatgaatt gtaaccttgg acaatactga ggcagcgtt caaagacttt ttggtctatt aggcgaggag
2801   gcagaacgtg acgggttgca agatactcca ttccgttttg ttaaaagcact cgtcgaacat acgtcagggg
2871   atcgagaaga ccttaaaactt catctcgaaa aaacatttga cgtcagacct gaagaccttg ttctgtgtaa
2941   agacattcca ttcattcttt tatgtgagca tcatttagct cgttctgtag ggaaggtgca tattgcatac
3011   attcctaagg ataagattac aggtctttca aaattcgggtc gagtgggtga aggatagcgt aaacgacttc
3081   aagtacaaga gcgcttgact caacaaatcg ctgacgctat tcaggaaggt cttaaatcttc aagcagttgc
3151   ggtcatcgta gaggttgagc atacttgcat gagcggagcg ggtattaaag agcacggggc aacgcagctg
3221   acttcaacta tgcgaggtct tttccaagat gacgcactct ctcgagcaga attgtctcag ttgattaaaa
3291   agtaggaggg ggaataatgaa taaaagtcca accttttggc ttgttcgaac agctcttatt gcggctctat
3361   atgtgacatt gaccgttgca ttttctgcta ttagtattgg acctattcaa tttagagtc a gtgaagcctt
3431   gattcttcta ccttttatgga accatagatg gactccgggg attgtattag gaacaaattt tgcaaacctt
3501   ttttccacct ttggaactga tgacgtttta ttccgttcaa ttgctacctt ccttggagta gtggcaatgg
3571   tgaaagtgtc taagatggca agtcccttat attcaacttt ctgtccagtt cttgtctaag cttaccttat
3641   tgcgtcggaa cttcgaatag ttactctttt accttttttg gaactctgtc tctatgtagg aatttagtga
3711   gcgattatcg ttttaatttc atacttcttt atttccacgc tggcgaagaa caatcatttt agaactgga
3781   taggagcgaa aaatgggatt caatctatac ttccgaggag gtcacgctat tagcactgac gattatttga
3851   aggaagagg agccaatcgc ctattcaatc aactgtacga aagaaacggg atttggcaaaa ggtggattga
3921   gcataagaaa accaatccaa gcaactactc aaaactattc gtcgactcta gtgcatttc tgctcatacc
3991   aaaggggctg aagttgacat tgacgcctat atcgaatacg tgaatgataa cgtgggaatg tttgactgta
4061   tcgcccgaact cgataaaatt cctggtgtat ttgacagacc taagacacgt gaacagcttt tgggaagcacc
4131   acaaatttct tgggataatt atctatatac gcgcgagcga atggttgaga aagacaagct cttacctatt
4201   ttcctatagg gagaagactt taaatggctc aacttgatgc tcgaaactac atttcaaggc ggaagcaca
4271   ttccttacct tggaaattca ccagccaatg actcgactac gaagcataaa gacaagtgga tggaaagagt
4341   attcgaagtt attcgaacaa gttctaattc agacgttaag actcacgctt ttgggatgac agttacttagc
4411   caatttagag gtcacccatt ctatagcggc gactctactt ctgtactgct cacagggagc atgggaacaa
4481   ttatgacgtc aaaaggatga gttgacttgt cacagaagaa tggaggaatt gatgctgtcc gtaggctgcc
4551   aaaaacgggt caagttgaaa ttgaatccat tatcgaagaa actggagcgc atttttagct agagcaatta
4621   gttgaggact ataaacttcg agcatttgtt aatgttcaat acatgctgaa ttgggcagag aactatgaa
4691   tcaagggaat taaaactcgt caacgtcgac tattttagat aagagctttt cgtctctatt ttttttaaaa

```

4761	aaaaatgaac	tttttatata	aaaacgcttg	actttattca	ctcattatcg	tataatcata	atataataaa
4831	aacgaataag	aggtaaataa	aatgacagca	gttcaacaag	ttaagttcta	cttagaagaa	gccggcgctc
4901	actttctaaa	agatgttgag	tacagtgcac	acttagagca	agcaattatg	aaagataatc	tcaaatggaa
4971	tggcgctcat	agagatgagc	acgatcatgaa	aataaacttca	tacgaagtat	taragagagg	ggtaaggcta
5041	tgaaaaaagt	tcaaaacttat	caagaatatac	taaaactagt	tgagtccaac	cgtaaacctt	ctttaaactt
5111	tcgagaagga	aaaaataggag	tcgatgaagc	ggttattcaa	ttattcacct	tctatagttt	caacaatatc
5181	gaggaacctc	ctttcattgt	actcaaaatg	caagaggctg	ccgtgaacgg	gacttatgaa	gcaaaactca
5251	atatgcttaa	aagattttaa	attattttaga	aacggcttta	caaaactcgcg	ataattcgtg	tatatatat
5321	atatcaaaaa	aaggaggctc	atattatgag	tattcaagttc	aaaaccgaag	aactttcaaa	aattgtttct
5391	cagctcaata	agttgaagcc	tagcaagtgc	ctagaaatca	caaaactattg	gcataatttt	gggtgacggc
5461	aatgctcat	gtttacacgc	tatgatggct	caaaacttctc	tcgatgcatt	atcgacagcg	atgttgaat
5531	tcgctgatt	gtgaaagcag	agcagtttgg	aaaacttgta	gaaaagacca	cgcccgcaac	cgtaacatta
5601	gttccctgaag	aatcttcgct	aaaagtattt	gggaattggtg	agtaacaata	tgatattggt	acagaagatg
5671	aagagratccc	tacatttcgac	cacttgctcg	aagacgtgag	tgaagaaat	gctctcactt	tgaagagctc
5741	gctgttctac	ggaatcgcca	atatcaacga	ttctgcggtg	tctaaatcag	gagcagatgg	aatttatacc
5811	ggcttctctg	taaaaggcgg	aaaagcaact	actacagaca	tcatttcgctg	atgtatcaac	ccatcaagg
5881	aaaagggaact	agaaaatgctc	attccttaca	acctaagtag	tatttttagca	agtatttctg	atgagaagat
5951	gtactttctgg	caaattgacg	atactactgt	ctatatattca	tcggcttcag	tcgaaattta	tggaaaattg
6021	atggaaggta	tggaaagatta	tgaagacgtt	tcacagcttg	actcaattga	gtttgaagat	gatcgcgcta
6091	tccttcacgc	agaaaactctg	agcgtattag	accgcttctg	actattcact	tcagcctttg	acaaaggaaac
6161	cgctgaattc	ttattcttga	aagaccgact	tcgaattaaa	acttctacta	gcagttatga	agacatcatg
6231	tacgcatctg	ctggcaagaa	agtttctgaag	aaagaattca	cttgccacct	taacagctca	ctcttgaagg
6301	aaattgtatc	aacgtctacc	gaagaaaact	tcactgtctc	ttatgggaagc	gaaaccgcaa	ttaagatttc
6371	atcgaaatgg	gtcgtttact	tcctagcact	tcaagagccg	gaagaataat	ggccaagtcc	aatttaacta
6441	gaattgcaaa	gatggttaga	gcaggaaaca	gtgaaggctc	tgcttcatct	tttgtcaatt	cgctgacccg
6511	ggttattgaa	cgaactcagc	ctgaatatata	tccttcgaca	tattataaagc	ccagcggggt	tgggtgattg
6581	attcgaaaaa	tgtatttctga	aagaatcgggt	gagctctatta	tagataaacg	agatttcaac	ctaattgcaa
6651	tgggctgaagc	tggaaacattt	aggcacgaaag	ttctccaaga	gtacatgggt	aaaatggcgt	aaatcgatga
6721	ggactttgaa	tggttgaatg	tagcagagtt	cttgaaaagaa	aatccaagt	aaggaaactat	cgctgacgag
6791	cgtttcaaga	aaaacagatta	tgaacagag	tgtaaagacg	aacttcttca	actttctatc	tgtgtgagc
6861	gactagttcg	atataaaggc	aagctctaca	tttttagagat	taagactgaa	accatgttca	agttcactaa
6931	acatactgag	ccctatgaag	aacacaagat	gcaagcaact	tgctacggaa	tgtgtctagg	agtcgatgat
7001	gtcattttcc	tttatgaaaa	tcgagataac	ttcgaaaaaga	aagcctacac	gtttcacatc	acagacgaga
7071	tgaaaaaatca	agrccttgga	aaaattatga	ccctcggaaga	gtatgtagag	aaaggcgaaa	gtccttaaat
7141	ctattgtctc	tcagcctatt	gcccatattg	tagaaaaggaa	ggctgaaatc	tgtgagctat	actggaaaaa
7211	tgttcgagga	agacttttct	gaaggtgcaa	aagactttga	gaaagatgct	ttcacgctcc	gtctatcatga
7281	taccactaat	ggatttcgag	gagttgcasa	tccttgcgat	tatatagcgc	caacraactt	tgggaccttg
7351	tttattgaac	tgaaaaactac	taaagaagct	tccttgagct	ttataaacat	cactgataat	caatgggtcc
7421	agctatcacg	cgcagatgga	tgcaaaattta	ttctcgccgg	aatttttagtg	tatttccaaa	agcatgaaaa
7491	gatttatatg	tatccaattt	caagccttga	aaaaattaaa	cggtctggag	ttaaaagcgt	caacccaac
7561	ttcatcgatg	cagggtatga	agtttcttac	aagaagcgctc	gaactagatt	gaccttccr	ttccaaaatg
7631	ttctagatgc	agtttgagct	ctttacaagg	agaaaagcaa	tggcaagacc	taagttacct	caaatgtgata
7701	ttcgagaaga	agaaatcaga	gatgtctcaag	acgtagcaga	ctcgatgggt	gcgattatca	ataaagtagt
7771	cgacgaaat	gttgaagcag	cttgcgggtc	acttgaccag	gcaatggaaag	aaattcaaat	agttgtgaagc
7841	caaaatccctg	tcattatgga	agaccttaac	tactacattg	gctatcttcc	cactcttctt	tatttcgccg
7911	cagatagggc	ggaaatgggt	ggaatacaaa	tggattcgaag	ttctgctatc	aggaagaaaa	aatacagtaa
7981	ttctatacatt	ttagcccgccg	ggaaaactat	tcctgacaa	caagcagaaa	ctcgaaaact	tgtcatgaat
8051	gaagaagtca	tcgaaaaatgc	ttacaagcga	gcctacaaga	aagttcaatt	aaagctagaa	caggccgata
8121	aggtattagc	atccttcaaaa	cgaattcaaaa	cctggcaact	agcagagttg	gaaactcagt	caaatatttc
8191	aaaaggagta	ttatttaaatg	caaaaagacg	tagacgtgaa	aatgatggac	cctaaacttg	accgattaaa
8261	atacacaggt	gattgggttg	atgtacgaat	tagttctatc	actaaaaattg	acgcccagag	cgccgatgtc
8331	tcaagatgct	gaaaagtgct	tcaaaaggct	caagtatact	cagtgggcgc	aggtgaatgc	actaaaattg
8401	cacacggatt	tgtctctgaa	cttccctaagg	gatatgaagc	aatcttgcat	cctcggtcca	gtctttttaa
8471	gaaaaactgg	ctaatctctg	ttcttagcgg	agtgattgac	gaaggttaca	aagtgacac	tgtatgaatg
8541	ttctcagctt	ggtagcttac	tcgtgacgca	gatattctct	acgaccaaag	aattgcccac	tttagaattc
8611	aggaagaa	acctgctatc	aagttcaatt	tcgtagaatc	tttaggaaat	ggcgctcgtg	gaaggcatgg
8681	aagtacaggt	gattttctaat	gaaattggaa	cagttgatga	aggactggaa	taaggattcg	aaagctcttg
8751	tagcagttca	aggacttgaa	cgtgaagcgc	ttccaagaat	ccctttttct	gcgcttcta	tgaattatca
8821	aactcagcgc	gggctccctc	gaaaaagggt	agttgaattc	ttcggctcctg	agtcgaagtg	gaaaaactact
8891	tcagctctcg	acattgtcaa	gaatgcgcaa	atggtatttg	agcaggaatg	ggaacagaa	actgaagaa
8961	tcaaggaaaa	gctggaaaaat	gcgctgcat	ccaaagctag	caagactgct	gtcaaggaa	ttgaatgca
9031	actcgatagt	cttcaagagc	ctcttaagat	tgtatatctt	gaccttgaga	atacattaga	cactgagtg
9101	gctaaaaaga	ttggagtcga	tgttgacaat	atttggatag	ttcgccctga	aattgaacagc	gctgaagaaa
9171	tacttcaata	tgttttagac	attttcgaaa	caggtgaaat	tggcctagta	gttctagatt	ccttgcttta
9241	catggctcagt	caaaacctta	ttgatgaaga	gttgactaaa	aaggcctatg	caggaaattc	agcgcttttg
9311	actgaattta	gtcgaaaagg	tactctctct	cttactcgct	acaaatgcaat	attcctagcg	atcaatcaaa
9381	ttcgagaaga	tatgaatagt	cagtacaatg	ccatttcaac	tcaggcgga	aagatgtgga	agcatgtctg
9451	tgcagtctga	cttaaattta	gaaaagggtga	ctaccttgac	gaaaacgggtg	catcattgac	ccgtactcgt
9521	cgaaaacctg	caggggaatgt	agtagagtca	ttcgtcgaga	agaccaaaagc	attttaagccg	gacagaaaat
9591	tagtttctca	tacgctttcc	tatcatgatg	gaattcaaat	tgaaaatgac	cttgtagatg	tcgctgtcga
9661	atttggagtc	attcaaaagg	caggggcatg	gttcagatc	gtcgaccttg	aaactggaga	aattatgaca
9731	gatgaagacg	aagaaccatt	gaagttccaa	ggcaaggcaa	atctagttcg	acgcttcaag	gaggtgact
9801	acttatctga	caggtgatg	actgcgggtc	acgaaatcat	cactcgagaa	gaaggctaat	gcaaaaactc
9871	ctattttggc	ctaagctagt	gcctgctagt	tcaaggcgca	agaaaagaa	ggttccaaaa	cctaacccta
9941	aaatcgatga	gcaagtggtt	gagcttatga	accgcagaga	gcgtcaagtg	cttggtcata	gttgactcta
10011	ttattatttt	aatgactcaa	ttatagcaga	cgggcagtat	gacaaatgga	gccacgaact	atatctctct

10081 atagttttcgc accctgatga gtttcgacag actgtttctct ataacagagtt taaacagttt gacggaaata
 10151 ctggaaatggg tcttccatac gactgtcagt ttgctgttaag ggtcgcagaa aggccttttaa gaaaatgaat
 10221 ttagcttcta aataccgtcc tcaaaccttcc gaggaagtgg tagctcaaga atatgtcaaa gaaattctct
 10291 tgaatcaact acaaaatggc gctatcaaac acggtatctt attctgtggt ggcgctggaa ctggtaaaac
 10361 cactactgct cgaattttcg cgaaggatgt gaacaaagga cttggctctc ctactgaaat tgatgtgct
 10431 tctaataatg gggtagaaaa tgttcgaaac attattgaag attctagata caagtctatg gacagcgagt
 10501 tcaaaagtta catcattgac gaggttcata tgccttcaac cggagcattt aatgcgctgt tgaaaacatt
 10571 agaagagccc tcatcgggaa cctggttcat tctatgtact actgacctcc aaaaagattcc tgacactatt
 10641 ctcagctcgag ttcaacgggt tgaactttact cgaattgata atgacgacat cgttaartcaa cttcaattta
 10711 ttatcgaaag tgaatatgaa gaaggagctg gttatagtta tgagcgtgac gccctttcgt ttattgggaa
 10781 acttgcgaat ggaggaaatc gtgacagtat cacaaggctc gaaaaggtcc ttgattatag tcatcagct
 10851 gacatggaag cctgtttctaa tgcactagga gttccggact acgaaacatt cgcttcaact gttgaagcta
 10921 ttgccaacta tgacggctca aagtgtttag aaattgttaa tgactttcac tactcaggaa aagacttgaa
 10991 atttagtact cgaactttta cagacttctt tttagagggt ttgaagtact ggctagttcg agatatttca
 11061 atcactcaac ttctgtctca ttgtgaaggt aagctagagc aattctgtga ggcttttcaa tctctactc
 11131 tattgtggat gctagaagaa atgaatgaac ttgctggagt tgttaaatgg gagcctaagt ctaaacccat
 11201 aattgaaacc aaacttcttt tgatgagcaa ggaggagtga catgattgga cagggacttg ttaattctac
 11271 catctcgaaa tggaaacaac ttccaaaata tataatcgct gaaggtgaag taggttcagg acggaagacc
 11341 ttaactcgtt atattgcttc gaatttgac gctgattcta ttgtagtagg aacgagtgtg gatgacattc
 11411 gaaacatcat tcaggatgca cagactattt tcaaggcgag aatctacgtg atagacggaa atagcctgtc
 11481 aatgtcagct cttaactcgc ttgtgaagat agcggaaagc ccacttttaa actgtcatat agcctagact
 11551 gttgatagca tcaataatgc tttaacctacg cttgcaagta gagcaaaagt tctaacctatg ctacctata
 11621 ctaatgaaga gaaaatgcag ttgtcgaagt cctacaagaa ggtagatact tcaggaaatg acgaccgagc
 11691 taattgtagac tattgcaatc ttgcccagaa tcttcaaatg cttgaagaca tartaagaata tggcgagaa
 11761 gagctatttg aaaaggttac aacattttat gacttaatat gggaggcaag tgcctagcaat tcgctaaagg
 11831 tactaattg gctcaaattt aaggaaactg atgaaggaaa aattgagcct aaacttttcc tcaactgtct
 11901 ttraaattgg tcgacagttg tcatcaggaa gcactatgta gaaatgtctt tcgaagaact tgaggcccat
 11971 gaccccttag tgagggaagc atctaggtgt ttgcaaaagg tatctaaaaa gggctcaaat gcgcgtgtct
 12041 ggtggaacga atttcatcag aggtcaaac aagttgagtg atttagtatt atttcaaaa gacattcgaa
 12111 ccaataatct aaagccgttc tatatcttgc acggcgaaga aattggctct atgaatgttt atctaatca
 12181 aatgggaat gtatgtcgag aaacttcggt ttcaacagtc tggaaaacc tcaactcaaaa agggctcgtt
 12251 tctaactcgc gaatatctgc ttttcgagat gataaggagt ttctgtctaa tgagtcgagg tggaaaaggc
 12321 tcccggtagt tagatatggg acacttgttt tgatgggtac taaaattgac aagcgaagca agttgtctaa
 12391 ggcctttcct gataattgtg ttgagtttga gaaaatgact gacgcgcagt tgaaggca ttttgtgtct
 12461 aaatactcga cttatgatag cgacatgatt gacttggtta tccagttctg tctaaccgat tactctagaa
 12531 ttgacaatga attggaacag ctgtcgcgat tgaaaaagg ttgacgcatca gtatgtgaat ccattgtcaa
 12601 gcacaagacc gaaattgaca ttctcagcct agttgatgat gtatttgaat ataggccgga gcaggcaatt
 12671 atgaaagtga ctgaactttt agccaaggga gaaagtctta ttggattgct taccttgctt tatcaaaatt
 12741 ttaataacgc ttgtcttctg cttagggccg atgagcctaa agagccattt ctaggcatta agcagttctt
 12811 aatcaataag attgtctata actttcaata cgagctggac ttagcctttg aaggcatggc tatcttaggt
 12881 caagctatcg agggcataaa gaatgggtgc tatacagaaa gttcagttgt ctatatttct ttgtataaaa
 12951 ttttttctact tacttaacaa ataagctgaa atctgtgtat attacagtat aagcaaaagg ggcagacctt
 13021 tgacagaagt tgcggttaaa agcccgcaaa aggtgagagt agttatgggtc gggaaatctg aatttctcga
 13091 atatttaaaa aggaagctag gaacagaaac ttccatcagt tatattatag aaaaatgaag gggctcaatc
 13161 tgacagactt taaaaaacgc ttcaagaaag cagtaacaga aacaatcaat cgtgacggtc tcgagaacct
 13231 tatggattgg ctggaataatg ataccaattt cttctcaagt ccagcaagca ctcgatacca tggaaagctat
 13301 gaaggtggac ttgtcgagca ctcatcaac gtgttcaatc aactactttt cgaatggat accatgggat
 13371 gcaagggtcg ggaagacatt taccatggc aaacagttgc aatcgtagca ctatttcacg accttgcga
 13441 agttggtcag tatcgtgaaa ctgaaaaatg gcgcaagaac agcgacggtg aatgggaaag ctatttagca
 13511 atgaaatcag accctgagca acttcaaatg ggacatgggt caaaatctaa ttctctctt caacttctca
 13581 tccaactcag ccagattgaa gctcaagcaa ttttctggca tatgggagcc tatgatatta gtccttagc
 13651 aaatttgaat ggaatgtgag cagccttcca aactaatcca cttgcattct taatccatcg cgcagatag
 13721 gccgcaact atgtagtcca aaatgaaac ttcgaaatct ctcaaggtcc agttgaacaa gaggctgagg
 13791 ttgaagaagt agttgaagaa aaacctaaaga gttcaactcg taagaaacct gcgcctaagg aagaaaagc
 13861 tgaagaggtc gaagaaaac caaaagctgg aatcactcga cgtcgcaaac ctgcccacaa agagggaag
 13931 gtagaagagc ctaagaaga gcctaagaa gcattctcta aaatcgaat gcttaaaaag actgaaaagg
 14001 tcgaagaggt agaaaagcga gacgagccga aagttgaaga agcagaggac gacaatgtgg tggtaacctg
 14071 tggatagtgt cgagatgtct actacttcta cagtgaagtc gctgacgtrt actacaagaa agatgtcgac
 14141 gagcctgacg atgacagcga cattcttcta gacgaagaag agtacctgga cgcaatgtgt cctgtattag
 14211 aagaagactt cttctacgaa cttgacggca aggttcacaa attagcaaaa ggtgaacgct tgcgggaaga
 14281 atcagacgaa gaaacttggg aacctatcac tgaagcagaa tacatcaagc gaacsgaaaa acctaaagca
 14351 gttgcataaac ctactcgaaa aactccagcg ccttctcgtc gccctcgccc ttaaaagaaa ggttgaata
 14421 aaatgtgtga aaattgtcaa aacgaacat tcaatactag aattttcaat gaagatgaaa gtggctatgt
 14491 cagcgcctca ttcacttaca aggagattcg cgacaccgca gcagctatta gcaatcgagc ggttagaaa
 14561 aaagaccgtg acagcctttt agtcgctaca gttatggctc ttccgcttc tcacgcagaa gatttaggca
 14631 agagactttg tattgcaaat tctcgattgg aagcatttct tgaagctgtt caagaggctc tcgagaatga
 14701 aaaggctgaa gatttaaaag acgttatctt aggtcttctc gacgttgaca aaaaaattgg caacttgcga
 14771 ttgcaattag ttgaatcagg agcattataa tggaaacgaat aaagacgcta tttcacgtga tttatgctaa
 14841 cggcactcat ttgaagtaga cagcttttgt cgataccgtt gatgattatg atgacgttat agaggacac
 14911 caggggtgata ttgatcccc tgacctttat aatcaaaaga gcattagaat ggccgcttac aactctgaca
 14981 tcaactggta cgtatttgc actgacattt tactacgact agatgatatt atctacgtcg acgcaacttg
 15051 tgaactattt aaatacagag agcctatttg atgaacaatc agcgaagaca aatgaacaaa cgaactcgtg
 15121 aacttcgaga agactatcaa cgtgcaagag gtcgaataaa cttccttctt gctgtaagg gacacggcga
 15191 agaaactcga aaccttgaag ccttctgtgg atacattgac aatctagtcg aatgttttcc tgaagccaaa
 15261 cgaatgtct tgaggctatg tgtattagat gaccttccag tcaactatgc ggccgctgaa attggaatcc
 15331 actatacatg ggttaccacaa cttcgagaca aagcagttga aacacttga gaaatttag atgggataa

15401 cattatttcgc tctaaacacg gaatcgaaat taaggagaaa cttgatgaat tatatggtaa aagtcattct
 15471 agttagtgtc ttgtactgt cagccttttg catgacttgc tcaatgggtt atttgggtac aggttaagcaa
 15541 gagaccacc gtagtaccgt cgccttctga ttggcgctc tctgaagctc tgccggcttc tattcgacac
 15611 tctttatcct cgctatctg ccacgacatc acgcgcatac aaaccaatcc ccacgcgcag agctagtgtc
 15681 aaacaagaga aggcagttgc taagcagttg ggagggaaaag tacagcctaa ttcaggagcc actgactact
 15751 acaaaaggtga cgcgtaaca gactcaatgc ttatagaatg caagacagtt atgaagccac aaagtccagt
 15821 cagcttgaaa aaggaatggt tcttaaaaaa tgaacaggaa aggttcgctc aaaaactcga ctattctgct
 15891 atcgctttcg accttgggtga cggaggcgaa cagtatatag caatgtctat aagtcagttc aagcgaatat
 15961 tagaggatag aaatgataac cttattttaa ataaacagtg aaggaacagt tactccaatt aaagggctcag
 16031 ccacgcaact gtacgcagac cttattccta tacaagagga cgatatacac ttcgttgata taactggact
 16101 tgaccttatt gttcgagaaa acgtacttga gctcatttca cggagccgtg taggagcttc aaaatatggt
 16171 acaaacctcg accagatga tgcgcagat ttctacagc acgccaaga agaagcgctc gactttgtca
 16241 actacctaac caagctacaa agtcaacaaa agcaaaataa atagacctat ttctaggctc atttttatta
 16311 ttgataaact ccagcaattt gacgagcgca atcttctagc gcagatacta ggtggcggtc ttctgttta
 16381 ccttgttcat ttctgtctt aattcttctg ttaaggcggt cgattcttct agttaatttc ttgatgattt
 16451 caattctcagc atcaacttcc atgtcgcgag taagtgtgac tccagtttca gcgacaggac atgctttgaa
 16521 tactgcaatg tcaagttcgc tcttttctaat aactgagcct aggtctaagt acaagttagg attgattcca
 16591 gtgaccttat attgtttctc agtttctttt acaggaatgc ttctatagtg gaaagtgtag ttctgtgtac
 16661 cgtctttcca atctgtctga agataaacga aataaagtgt tgtttccata attgacctc ttctgcgtcc
 16731 ttgacgcttg ttttatttat attatgatta tacgataata aaggaataaa gtcacgact ttctacaaaa
 16801 aagttgaact tttttaaata tttttttttg aaaaataaaa gccctaataa tagagctttt agtttagcag
 16871 aaaaattagt tcatcttcat aagcaagaat ctgcccgtac cggtaagaaa tagctgattc aatatccggc
 16941 atttcgttga ctcctttttt aagttcgtcg atagacagc tacaatgacc tattcttgac tgaagttcct
 17011 caatccttcc gagtctgttt tcattttctg tatcaattgt ttctgagctc aggtgagtgaa aggaacttgc
 17081 aatagtttga atggcttcaa aaaagtcctg tatgaaact cctttataag aaagctcatt cctgtgtatag
 17151 caggaaagca aagcgttcca gctagtgatt tgaatttgag ggttaggaga gtttcgataa cgtacaaaat
 17221 ttagaataac tttgtagtca atacagcct cagttatgatt gttgataaat accttctatt tataacctt
 17291 ccaaatcttc gtcctcgtca tgcgtttcat agcaggcgat aacttcaacc cactcgtctg cctcaccttc
 17361 gtttcgaact cgaatgctaa ggacttccat gtcctcaaca tcttcgaatc cttcattagg tctcatctc
 17431 tcccactcta aatcgtcgtg gtcgaagata gttacaagac gtccgctcaa ttttactgtt tcccttactg
 17501 ttgctcattt agtttctcc ttatgcgata tatagtttga taatttgaga ttcgatgtca ccatagttga
 17571 tgaacttaac ttgggtcgacc gtttcttcca tgtattcgcc catgtcttgc atttctcgtt cttgaatcat
 17641 ttggccgttt tegtgtataa ttctgtacca ccattcatca ccgaattgtt tgattgtctc tttaactggt
 17711 ttcatcttcc tacctccact ttttcttcca ttagtgttcc gttatcatag aacgcaatcc gtcctactc
 17781 aagacgttct aggtcttacc atttaccgac ttgacgggtc gttactttaa attcagttac ttttgactt
 17851 caaactttca tctctacttg caaatcttta acttttacca ttttatatga ctcctttatt tgttttctt
 17921 tatagtatta ttatcagata atgagtgaa aaagtcaagt gtttttgaat acttttttaa atttttcaat
 17991 tttttttttt aaaaaataaa cgagccgaag ctacgttatt tatttatctg ctcaagggtc tgttgaattg
 18061 cctcatagcc ttacagcgt gctaccttcc cagctttaga gccgggtgaa aagtcacaaa cagttctgct
 18131 tactttaaag tcatcgcctc tggcatagtc gacgaggagc tggatagctt ttggccattt ccgccaaatc
 18201 ttggaaaact cacttatatt agcaccaacg aaaaacagtg ctctagtatg ctggctagac ataataaact
 18271 ctaaaaagtt gtccaaggtt ataggaaggt ccttgggaaa ctcataaggc tctttgacat cgtatttgaa
 18341 aaggctgaca atttcactgt ccttaaatag ttcaccgtct ttatcacata taccttgaac atttctagta
 18411 ggctctgtcc cgctatctag tacatcgcca accgtgtgac aataggcttt aagaacttga aaaaacctg
 18481 gggcgtctgc acgcgcaacc tggagcttcc taacagtcac ccaaggctga ggtttcttac aaacaactcct
 18551 aattccttca aaatagctct tgcctgggtc aatagtgcct aacattgtca gcctgttttt atttatataa
 18621 aggtcgaaat atacttgaat ttcatctgta ttaggcagcc acttaacagt gacttttcta taagcgattg
 18691 cttttacatt tactttttt gagagatttg tagggataag ctttttctt ttgacattta ttttttctg
 18761 ctttttgttc ttggcctatg tagtatctcc atttctgtg gcttctgtt ttagctctgt tcagttcagc
 18831 tgcttctcgc gatgcaatag ttctgagaat atgcctgttc ataggctcac aatattccgc caaagatttg
 18901 ccagttatgg tggcgtcaat taagttaacca tctattgact ccttaccata aaatacaaaa cgtcttggc
 18971 atactagcct ttataatag ccatcttctg cgcgtgttcc aatttttaact aagctcattt tcaaccaaac
 19041 ttgtagacga taaggagttc ctggaacttc gaacaggagc ctcctttttt catcgtctac ttgtttaata
 19111 catgagtttt gaaaatggat aactttccat ttattttcca tagtttccac ttattccatg taccogtcaa
 19181 caatccataa ttgaaaaggc ttatcttctc tataaggccg tgataatttt agtccagttc ccactacatt
 19251 tgaagcgcg attaggtcat ctaggctgtc tagctcagc tcatatacaa ggttgccagt atcaatttca
 19321 caaaagtaag cgacatttcc aactttctct agtgcttccac gatacctatc atatgtcgc ccttctgcaa
 19391 atagtccgcg agaataaact tcaaatcca ttttagttac cgccttccaa aatttcatg ggcataactc
 19461 ttgcattctc gccatgaac cgccttccaa tatcgccttc aagattgaag tcatgttgag gtcgtcaat
 19531 tcttctcttc tttcaatttc gaaatgtgtc ctgaagcgca tttttgttt gctcgttagg taggaccata
 19601 agtgaatatt cttccacctg ctttttaaat cgaatggcta aggtcgacaa aaagcctttg aggtatgaat
 19671 tcttttagga aggttcgcga gtaggaggtc ggtcaatagc gtaacgaaga taaagcaaaag cagctcata
 19741 tatttttagac actaatccag cgccttgttt ttgcgcgaag aaaattattc gacttttatt caagcgcata
 19811 tcacgctgat taatacaaaa gcacctaaaa ttagtccgca gaatatgacc aagttcacgt tcccacaaa
 19881 atattccacc tgccttcttc ccaacagctc gagaagtttc gaactgttta ggttcatcaa attgttcaac
 19951 ttgagcaagt gcgataatat tcttttagcat caacttttga gccataagaa gggcagtttg cccctcttgc
 20021 tcaactcgggt tgcattttgc taattgaata agatttttaa ttttttcaat aatttttttg ttattcatat
 20091 tagtcacttt ctatcatatt tctagacttt cgaaggttca atgtcgtcta cttcaattgt cttgtcataa
 20161 gtccaagcgc gacaagtgtc gaaatgaaat aggtcacaaa acatcttttc atttatggctg aaactttcag
 20231 tacatttttc aatattctact tcaagtcca gaacgcaaat agtatcaaca tttcgaagcg ataaaaaggc
 20301 tagagccttt tcaataacttt ctgctaggta aataactcca gctgaaggct tcaatcttc agttgcaatt
 20371 ttaccaagat tatcaaaatc agtggcggtg taaagtcca ttagttactt cttacatat cttaggtcac
 20441 tacataataa gaagcagttt tacttcccaa gtcctactca atagcttctc cttcgtctgag ttttctcagt
 20511 tttaaaactg tgcgttcagc tacaacatta gcaaggttcg aacggttgag aatgttttgc atatttctgc
 20581 cgcctaagac tttagcttgg tcatgtttca ctaccattag gtattcatta gtaagtgtt tagcaaggtt
 20651 cgaaaatttc attttatttc cctttatttc gttttcttct atactattat tatacaataa tgattgaata

20721 aagtaaagca ttttttataa aaaagttgaa ctttttttac aattttttga actattttaa aattataaaa
20791 tgggtggaaa atttaggcga caattttata cctattttac cctcatttat aaacaattcta atactagaaa
20861 ggacttaata agtaataaaa aaagcgccct gaaaaatcct acaaatccca tagtccgtaa gtaaaaaaaa
20931 aaattagggg cgacataaaa gtcgagcact atcttaatct attaccagtc tcatatataa tcgacacaga
21001 ttttagcaggc ttttagcaaa ctttcgaaca gcatgaaaaa gcatacaatt agaggaaacag attatagaaa
21071 aagcacttcc acaaaacaagt tctcaaaatg cttcaaaaaa ccgtaaaatt agtaagtttg aacttttcga
21141 acttctaaac ttttcgaata atcgagccta atttagaggt cgaaaaactc aattttctcga aaagtcgaac
21211 ctgctcgaaa acctcaaaac actcgaaaaa tcgagcatag aaaggggtcg. aaaagtcgag aatgctcgaa
21281 aaactcaacc ggttcgaaaa cctcaatcct tcgaaaaagtc gaaccattcg. aaaagttaa aagttcgaaa
21351 aactcaacca ttcgagagta ggaattaaag acataccagt tcaacctttt tagcttcaaa atcactcttt
21421 tttcatttat aggactataa attcagtc aa tcgtaagtca cgcgcaaat tgttacaatg taacgataa
21491 aatataaagg aggggtcaata aatggcgaaa gctactggac caaaagtrcg aagaggaaaa actcctccac
21561 gggcaaaaaga caaaaaagga atcaaaagcaa atgctgctgt caataaagac cagttcgtag agtatgacta
21631 taaaggcatc aagatgacaa ttaaggaaacg tgatgctaga atgaaattgg aattttattag agtcatgact
21701 attcaggaaa ttcgagcccg ctatggatta aatgaaaaagc gtgttggcga aatacgggct cgcgataaat
21771 ggggtgaaagg taagaaaagtg ttcgagaatg aaaaggctct tgttactaat gatacattga ctcaaatga
21841 tgcaggtgtt aaagtctcag tcaattatga atatcacgccc gcttgggaga aactaatgaa catgctcgaa
21911 atgtgtttag ataactctga cagatattta tttactaaag aaggaaatat tagatggggc gcattagatg
21981 tcccttcgaa ccttatagat agagctcaaa aaggacaaga aagagcgaat ggaatgcttc cggaagaggt
22051 tcgactatga ctacaaattg agcgcgagaa aattacattg ctccgggcca aatggggcga ccaggaaatt
22121 gaaggcgagg ttaagataaa cttcgtagaa gcactagata aagcagctca agccttttg caagaattta
22191 gtgacgcaac aggttccctac attaaaggag tgactgataa tgacaataag cctgagaaat aaactaccta
22261 agttcaactc ctcccttttt agtaagaaac aactccagct cctaaatagg tggacaaaag gctcacttt
22331 tcgaacttct gatattcgta tagcagacgg ttcattcgt tcaggaaaaa cagtatcgat ggctctttca
22401 ttttcccttt gggccatgac ggaattcaac ggacaaaact ttgccattcg tggtaagaca attcactcag
22471 ctgcagcaaa tggttattcag cctctaagc aatgctcac aagtgcgggg tatgaaattc gagatgttcg
22541 aaatgaaaaa ctacttatta tttagacatt tagaaatggc gaagaaattg tcaactactt ctatatatt
22611 ggaggaaaa atgagtcgag tcaagacct atacaggggg taacattagc aggtatcttc tgtgatgag
22681 tggcactgat gctgaaatcg tttgtcaacc aagcgacagg gcgctgttcc gtaaacaggt cgaaaatgtg
22751 gttctctgt aacccggcca atcctaata caacttcaag aagaactgga ttgacaaaac ggtcgaaaaa
22821 cgtatcttat atcttcaact tacaatggac gacaacctta gcttgacgga tagcattaa aggcgctatg
22891 agaaaaatga tggctggagtc tttaggaaaa gatttattct cggcctttgg gtaaacagcag atggcttagt
22961 ttattcaatg ttcaatgaag agcagcatgt caaaaagctc aatatagaat tcgaccgttt attcgtagca
23031 ggagactttg gtatctataa tgcaacaacc ttggccttt atggattctc gaaaagctac aagcgtacc
23101 actcaattga gtcatactac cactcagggc gcgagggcga agagcaacta actgaggcgg atgttaattc
23171 gaattattca tttagttcag tttcacaaaa gactactaaa gagtacycaa atgatttagt cgatagata
23241 cgaggaaaagc aaatcgaaata tataattctc gaccgtctg cttctgctat gatgttgaa cttcaaaagc
23311 atccttatat agctagaaag aatatcccta tctattcctgc tcgaaatgac gtgacgcttg gcaatttcatt
23381 tcacgctgaa ctcttgctg agaataagatt tacactcgac cctagcaaca cgcacgacat tgatgaatc
23451 tatgcttaca gctgggacag taaagcgagc caaacgggag aagatagagt cattaaagag catgaccac
23521 gcatggatag gaacagatat gctgttctca ctgacgctct aatcaacgat gacttcggtt tcgaaataca
23591 aatattatcc ggaagaggcg ctagaacta actaaacact tttatagaaa ttagtgtata atataagtag
23661 gaggatttta aacatggcta aaaaatcaaa agctatctca cacacagacg aactgattag tcagtcggtt
23731 gacagccctt tggcaaaaga tcaaaagttc aagaaagagc tttaggaagt tgaaaagtat tatcaatact
23801 tcgacggatt tgatgtcacg gacttgaata ctgactatgg gcaaacatgg aagattgacg aagactcagt
23871 cgactataaa cctactcgag aaattcgaaa ctatattcga caacttatca aaaagcaatc acgctttatg
23941 atgggttaaa agccagagct tatctttagt ccagttcaag acaatcaaga tgaacaggct gagaacaagc
24011 gtattctatt cgactctatt ttaaggaatt gtaaatctcg gagcaaaagt acaaatgcat tagtcgacg
24081 cacagtaggt aagcgggtat tgatgacagt agtagcaaat gccgctcaac aaattgacgt ccagttttat
24151 tcaatgcctc agttcaccta tacagttgac cctagaaacc cttccagctt gctttctgtt gacattgttt
24221 atcaggagca gcgtacaaaa ggaatgagca ctgaaaaaca actttggcat cattatagat atgaaatgaa
24291 agctggaaca agtcaatcag gaattgcaac agcttttagaa gacattgaag aacaatggtt gctcactat
24361 gccctaacgg atggagagtc gaacaaatc tatatgacag aaagtggcca aactactatc aaggagacag
24431 aggtcaaaact tgtagaattt gaagacaacc taggaaacaa gattgaagtt cctttaaaag ttcaagaatc
24501 cgcccccaacc ggcttgaagc aaattccttg tcgagttatt cttaatgaac cattgactaa tgacatatc
24571 gggacaagcg atgtcaaaag ccttatcaca gtacagata acttgaacaa aactattagt gacttacgag
24641 attcacttcg atttaaaatg ttcgagcagc ctgttatcat tgatggctct tctaagtcaa ttcaaggaa
24711 gaagatttgc ccaaacgctt ttgtcgacct taagagtgac cctacttctt caatcgccg tactggaggc
24781 aagcaagctc aagtcacttc catttcagga aacttcaact tcttccagc ggctgaatat tatttagagg
24851 gcgctaagaa agccatgtat gaactaatgg accagccaat gcctgaaaag gtacaggagg ccgcatcagg
24921 aattgcaatg cagttcttat tctacgacct aatctctcga tgtgacggaa aatggattga ggggatgat
24991 gctattcaat ggctcattca aatgctggaa gaaattttag caacagtga tgttgacttg ggaattatc
25061 ctcaagatat tcaatcaagt tatcaaacac ttacgacaat gactatcgaa caccactatc caattcctag
25131 cgatgaactt tctgtcaagc aacttgcgct cactgaagtt caaactaatg tacgagcca ccaatcttac
25201 atgaaagaat tcagtaagaa ggaagaggcg gacaaggaat ggaacgcat tttggaagaa cttgctcagc
25271 tgcaggaat ctacgctgga gcatgctg caattagcaaa cgaattaaac gaacaagagg agcctcaaga
25341 tgaaaacgag gaagaaagc aggttgatga caaagaaaaa gaacaaactg aacaacaaac cgaagaggga
25411 gtcgacccag acgttcaagg ttaattgtga ccattgtgag cataagttcg accttatact taacagatt
25481 atttcgaaac atatcgaaaa gggcgtagag tggagattct tcgaatgtcc taagtgcctt taccggtc
25551 ccacttatgt aggaacaaag gaaattgaaa accctattcg atttagaaat acttgcgag ctaaatgaa
25621 gcaggaaact caaaaaggag ctgctgctaa tcaaacact taccattcat atcgaaatca ggaatgacaa
25691 gctgggcata aaatctcagg gcttatggcg aagctaaga aggagataaa cattgaaaaa cgagaaaaag
25761 aatgggtatc tatatagctg ggaagaggct attcatgaaa ataattatcg tctaacctt gaacaggaa
25831 aagctgtact gaaagccttc agcgatgcag gaactgattt aattgcnaag attaaaaagt ctcgaaatg
25901 atacttgcct aaaaagaatc ataaagacta cgcttacgac ctgcacgctg tcttgttca actaatgact
25971 gaatactctc ataaggcggc aatgaacgca gtagatggc aggtagttca tattctacaa gtattagcag

26041 aagatggaaa tgctacggct gaaaagtctg aaaaggaagt cagggctgca tcttttagtat tttcacgaag
 26111 agcagccgag gcagttgtca aaggtgaaat ctataaggac ggcataaaacc tctcgaaacg tgtttggtct
 26181 tcagccgcac gccagggaaa tgatgttcaa caaatagtcac cacaaggcct agcaagtggc atgtctgcta
 26251 cagatatggc taaatgtctc gagaaatata cgcacccctaa ggctcgaaaa gattgggact ttgataagat
 26321 agctgagaag ctagggaaac ctgctgtctca taaatcatcaa aatctcgaat acaatgccct tcgacttgct
 26391 cgaactacca tttagccattc cgcacacagct ggagttagac aatggggcaa ggttaacctc tatgtctgaa
 26461 aagttcaatg gcattctgtt cagcgtccag gtcgaacgtg tcaagcgtgt atcgatttag atggtagaag
 26531 atttccatc gaagaatgtc ctttcgacca tccaatgga atgtgtacc aaactgcatg gtacgaaaaa
 26601 tcaactcgaag aaatcgctga tgagttaga ggctgggtag acggagaacc taatgatgta ttgacgaaat
 26671 ggtagcagca tttaagtcca ggaagaagtg agaatacag cgacctcgac tttgttaaaa gttattaggc
 26741 tcgggttcaat accgagtctt ttgtctcata aattgtctaa ttcgagaac cttcgaaaaag tagtaaaatg
 26811 atattcagtt atgttataat ataagttgaa aaggaacctt gtcgcttaa tgactcgaat ttggtttcac
 26881 tgttccaatt aaataaaaaa agcagattca gccggagggc ggaatactca ggaggaaaat aaatggctta
 26951 tcaattcga gacttgttaa aaggtcgaat tgaaccaact atcaaacagg tgaaggaaat tattcgaata
 27021 acttcgaaa aactcgatgc taaaattttc attgacggcg acggcacaac ttttgtacct cagcgacctt
 27091 tcgatgaagt tgttcaacag cgcgatgcag ctaacggctc aattaattct tataaagaac aatcgcgac
 27161 gcttctctaa caggtaaaag ataacggtga tgcgcagacc actatccaaa accctcaaga ccaactcgac
 27231 aagcagttct aacttgcaaa aggcgtctgt attacttcag ctcttcacct gttgattagt gactccattg
 27301 ctcagcagc agacattctt ggatttatga accttgacaa cattacggct gaaagtgcag gtaaaagttaa
 27371 aggtcttcat gaagagtga aagctgtctg tgagtctcgt aaactcttat tcaaaagaat ccaagttccc
 27441 gcagaacaag aggtcgaagc taagtccga gccgggactg gaaatttagg aaatccaggt cgtgtcgtg
 27511 gtgtgtgtcc cgaacctcgt gaaatcggct ctttttgtaa gcaacttgct gctgtcgaac aaacggcagg
 27581 agcacaagaa caatcatcat tctttaaata acaggaggaa ctaactatgc ctaattgtcg agttaagaaa
 27651 actgatttta atcaaacac tcgaagcatt gtcgcaattc ctgaccacta cgtgtctttg gctgtcga
 27721 ttcagctcac cgcagcaact caagtaggga acaagaataa cattcttgcc ggaacttgcc tgaaaaatgc
 27791 tactacattt gaaggacgca aaactggact cgaagtagta tctaccgggt aacaatttga cgtactatc
 27861 ttcgctgacc aagaagtgtt tgaaggtgaa gaaaaagtaa ccgtgacagt attagtccac ggaatcgta
 27931 aatatgcagc ctttcgaaaa gttggcgatg ctgtgcctga atctaaaaac gcaatgattc ttgtcgtaa
 28001 ataggaggaa ttatagatga atatttatga ttatatcaac gcaggggaga ttgctagcta cattcaagca
 28071 cttccttcaa acgctcttca ataccttga ccaactcttt tccctaatgc tcaacaaaca gggacagaca
 28141 tttcatggct caaggttgca aataatttgc cagtaactat ccagccatct aactacgacg cgaagcaag
 28211 tcttcgtgaa cgtgctggat ttagcaaaac agctactgag atggcatctt tccgtgagtc tatgcaact
 28281 ggtgaaaaag accgtcaaaa cttgcaaatg ctattgaacc aaagtctcagc tcttgcccaa ccaactatca
 28351 ctcaactcta taatgatact aagaaccttg tagacgggtg tgaagcgcaa gcagaataca tgcgtatgca
 28421 attgcttcaa tacggtaaat tcaactgcaa agcgaggctc aactacacta cgtactaac
 28491 atggatgcta agcaacaata tgcagtcact aagaatgga ctaaccacgc tgaagtgcac cctatcgctg
 28561 acattttagc agcaatggat gacatcgaaa atcgtagagg tgttcgacct actcgaatgg tcttgaaccg
 28631 aaacacttat aaccaaataa ctaagatga cttcatcaag aaagctcttg caattgggtg tcaaggttct
 28701 tgggaaaaact tcttgcctct tgcgaatgac gctgagaat tcatcgctga aaaaacaggt cttcaaatcg
 28771 ctgtctactc taagaaaatt gctcagttcg ctgacgctga caaacttctt gacgttggta acattcgta
 28841 gttcaacttg attgacgag gtaaaagtgt attgctcca cctgacgcag ttggtcacac cgtactaac
 28911 actactccag aagcattcga cttggcttca ggcggaacag acgctcaagt tcaagtctt tcaagcgacg
 28981 ctaccgttac aacttatctt gaaaaaatc cgttcaacat tgcacagtt gtaacagtt gtaacagtt
 29051 atcatctgaa ggaattgact atgtaggagt tctcacaact aattaggagg tgcgtatag tgcgtatg
 29121 aagctcttag cactttaatc gtttcggag cagtagtgca ttgagggtcg gttatttctt gccctgaagc
 29191 gcttgcctcg tctttaaatt aacgcaatt tgcgttcgag attaaggcgg ctgaagatgg agaaacggta
 29261 gaaactgttc tcaaaaataa tgaactcagtt gaagaaattg acgaagttag acaaatggcg gaagatag
 29331 cggctaaaaa cgttccctgag ctcgttgaat tagcaagagc taatggaatt gacatttctt caatttctg
 29401 aaaaagcgaa tatatcgag ctttaattaa gtacgaacta ggagagttaa atggcagctc aaacggacg
 29471 tgaatttagt aaaaatcaata tcgataacga taatttctccg tcaccaatga ctgaccaaag tatctcagct
 29541 ctttttagaca agcataaata tgcgcctat gttagtata tgatttgctt aatgaagacc cggaaatgacg
 29611 tggtaacctt tggacctatc agtctaaaag gtgacgcaga ctactggaaa caaattggcg aattctatta
 29681 tgaccaatat aagcaagaac agcttgaac tgatgaaaag tgcgaacgtg gttcgacaat cttaatgaaa
 29751 agggctgatg ggacatgagt tatgacgtga attatgttaa gaatcaagt cgtagagcca ttgaaaccgc
 29821 tcttactaaa atcaaggta cttcgaactc ttgggtcagt gatggatatg gaggaagaa aaaggataaa
 29891 gcgaatgaag tctgtagcga cgacctgtt tgttttagtg ataattcaac tgttctcgac cttttagcca
 29961 attctactga cgcgggaaaa atttttggcc aaatggagt gaaaattttc attctatag atgaaggcaa
 30031 aatcattcaa cgagcgata ctatcgaaat taaaaactca ggaagcaggt acagggtagt agaaaccac
 30101 aatcttctcg agcaagacat tttagatgaa cttaaatttg aggtgaacga ctaatgtctc agcctgaatt
 30171 agtatggaag cctgaagaat ttgttagtaa ctgtgaacgg tatcgaacaa agtttcaagt cgtgtcata
 30241 acagtctgag aagtcgctgc tactaagatg gaagaatag caaagacgca tgcattttgg acagaccgta
 30311 cagggaatgc tcgacagaaa cttcaaggag aagctgcttg ggtgaagcga gaccaaatca tgatagctgt
 30381 atcacatcac atggactacg ggttttggct agaactagct catggtcgaa aatacaaaa tctcgaacag
 30451 gctgtagaag acaatgtcga agaacttttt agagcgttga gaaggttatt agactaggag tgaactgac
 30521 taaacgaagc acaatgatgg acagattgaa ggaattctt cctacatttc agctctcgcc tgcctctatg
 30591 cttccaggag ttgaatttga ctagcaagat acagataggc cggatgacta cattgttctt cgtatagtc
 30661 atagaattgc cagcgcaaca aatagcctag gaagttttgc ttattggaaa gttcaaatc acgtccattc
 30731 aaactcaatt attggtatcg acgaatatag cagaaggtt cgaacattta tcaaggacat gggctacgaa
 30801 gtaacctatg cagaactcgg tgactacttc gacacaatgc tttctagata ccgactagaa atcgaaata
 30871 gaattccaca aggaggaac taatatgag taaagacatt cttacggaa tcaagctcgt gcaaaatcga
 30941 gagcttgacc cattgactca gttgccaana gtcggcgag ctaactttgt cgtagatacg gcgaacacag
 31011 cagaactcga agccgtgacc ttcggaggaa ctgaagatgt gaaacgcaat gacacgcgca tttctgctat
 31081 cgtgcttact ccagacctt tctacgggta tgacttaaca tcaaggaca acacgtttga ccttgaaatc
 31151 atggccctaa ttgaaggtgg tacagtacgt caacaaggcg gaactattgc tggatagcac accccaatgc
 31221 ttgcacaagg tgccttcaat atgaaacct ttagaatgaa catctatgtg ccaactatg taggtgactc
 31291 aattgtcaac tacgtgaaa tcaacttgaa ggttaagctc cagggctttc aatcgggaaa

31361 gagttctacg ctctgtaggt caacatcaag gcacgtgaag caaccaaaag aggtttgcca gttaaagtcaa
 31431 tggactatgt ggcacaaact ccagcgggtt ttcgtcgcgt gacatttcgt ttgaacgggt gaacaggaaac
 31501 cgccgacgca gttcaggttg aagcaggtaa gaagatttct ccaaaaccag ttgaccttac cttaaacaggt
 31571 aaggctttca aaggctggaa agttgaaggaa gaatcaacta ttggggactt cgacaaccac atgatgcctg
 31641 accgagacgt caaactcgra gcacaatttg catagaaatt tagaagaag gggtctgtat gactaatatt
 31711 atcacagctg agcagtttaa gcaacttgca ttcaaatca tgcacttccc aggtatttca aaaggtagtg
 31781 aacctatcca tgttaaaatt cgagcagcag gtcgatgaa cctaactcgt aacgggaaaa tccctaatac
 31851 gctttttagg aaagtgcag aactgtrtgg agaaacttcg acagtcacta aagacaatgc tagtctagca
 31921 tcaattactg accaacagaa gaaagaagcg ctgcaccgat tgaacaaaac cgataccggg attcaagaca
 31991 tggctgaact tcttcgagta ttgcgagaag cttcaatgg agagcctact tacgctgaag tcggcgagta
 32061 tatgacagat gagcaactta tgacaactct cagtgcgaat tacggtgaag tgactcaagc tgaacacctt
 32131 cgtacagacg aaggaaatgt ctaatgcat agcagtcgct actgaatttc atattagacc tagcgagtg
 32201 gtccggatgc aaactgattt aggcataatc tgcctcgcag cagcagccgt tgccttattt agatatttgc
 32271 aggaagacaa gactcctagg tatcctgggt acgaaaagaa aaatccagga ttgcaaatgc ttatggagt
 32341 actattttca gtcgtctccc tttttgtata tagaaaaggaa attacatgga ttttgggtca attgcagcaa
 32411 aaatgacttt ggaatattca aacttcacaa gtcaattaaa tcttgcctca agtcaagcgc aacggctcgc
 32481 actagagctc tcgaagctct ttcaaatgg ttctgcttta acaggtattg ggaagagact tacgactgcg
 32551 gttacccttc ctcttatggg atttgcagcc gcctctatta aagtagggaa tgaattccaa gctcaaatgt
 32621 ccggtgttca agctattgca ggagcgacag cggaaagagc ttgtagaatg aagactcaag caatcgacct
 32691 tggctgctaaa actgctttta gtgcaaaaga ggcggctcaa ggtatggaaa atctagcttc agccggtttc
 32761 caggtaaatg aaatcatgga cgtatgcga ggggtacttg acctggctgc cgtatctgga ggagatgtgg
 32831 ccgagagctc cgagggcatg gctagtctac ttgcagccct tggattagag gcaaacaggc cgggtcacgt
 32901 ggctgacgta tttgctcga cagcagctga tacgaacgca gaaactagcg acatggcaga ggcgatgaaa
 32971 tacgtcgcac ccgtgtgctc ctctatgggc ttgagccttg aagaaacggc tgcgtctatt gggattatgg
 33041 ccgagcggcg tattaaagggc tcgcaagccg gaaccacgct tagagggcct ctctcgcgta ttgcaaaacc
 33111 tacgaaagcg atggtcaaat caatgcagga attaggagtt tgcgtctacg acgcaaacgg aaacatgatt
 33181 ccactaagag aacaaactcg tcaactgaaa acagctactg caggactaac acaagaggaa cgaatcgtc
 33251 acctgtttac cttgtatggc caaaactcgt tgtcaggtat gcttgaccta ttagacgcag gtcctgagaa
 33321 attggataag atgacaaatg ctctcgtgaa ctgcgacgga gctgctaagg aaatggcaga aacratgag
 33391 gacaaccttg ctatgaaat cgagcaaatg gggaggagct tcgagttctg tgctattatt gttcaacaaa
 33461 tctttgagcc tgcacttgc taaaatcgtg gagcaatcac aaagttctc gaagcattcg taaatagtcc
 33531 acctatcggt caaaagatgg ttgtcatatt cgcaggaaatg gttgcagccc ttggaccact gcttctaatt
 33601 gtaggaatgg tgatgacaac tatgttcaag ttaagaattg ctattcagtt tttaggcca gcatattagg
 33671 gaacgatggg aacctatgca ggaattatag caatattcta tgcctgggtc gccgtgttca tgatcgttca
 33741 caaaaaatcg gagagattta gaaacttat caacagctct gcgctgcta ttaaagctgg gtttggagga
 33811 gcgttggatg ggctacttcc acgactgaaa gagttaggag aatggttaca gaaggcaggc gagaaggcga
 33881 aagagttcgg tcagttctga ggttcaaaag tgtcaaaact gctcgaaacg tttggaaataa gttcgttca
 33951 ggcaggaggc tcgattggct agttcatttg aaatggtctc gaaaggctag gaggcgcat ttgaaaagta
 34021 ggaggagctc tttcaattgc tgtttcactt gtaacaaaat tccggtctgc atttctaggg attacaggac
 34091 cactcgggat tgcattagt ctgttagttt catltttgac agcttgggct agaacagggt agttcaacgc
 34161 agacggaaat acctcaagtat tcgaaaactt gacaaacaca attcagtcga cggctgattt catctctcaa
 34231 taccttccag tctttgtcga aaaaggaact caaattttag ttaagattat tgaaggaatt gcattcgtg
 34301 ttcctcaagt agttgaagtg atttcacaa gatttgaata tattgtgat acaatttca cagttatgcc
 34371 tcaattagtc gaagcaggaa ttaagatact cgaagcgctt ataaatggct ttgttcaatc tcttctact
 34441 atcattcaag cagcgtgtca aattatcact gctttattca atggtctgt tcaggcactt cctactgta
 34511 ttcagcaggg tcttcaaat ttgtcagctc tcatataacg actagttcaa gcgcttccgg caattattca
 34581 agcagctgtt caaattatca tgcgtctgt tcaagcacta attgaaaact tgcctatgat aatcgagca
 34651 gcgatgcaga ttataatggg tctagtcaac gcaactgattg aaaatatagg acctattca gaaagcggga
 34721 ttcaaaattc aatggcttta atcgaggac ttattcaagt gcttctgaa ctaattacag cagcattca
 34791 acctattact tcaattatag aagcaactct gtcgaacctt cctcaacttc tagaagccgg agttaaattg
 34861 cttttatcac tcttcaagg gttgctaaat atgcttcttc aactaaattg aggggctttg caaatctga
 34931 tggcacttct taaagcagtt atcgacttcg tccctaaact tcttcaagca ggtgttcaac tcttaaggc
 35001 attgattcaa ggtattgctt cacttctcgg ctactttta tcgacagctg gaaacatgct ttcattcata
 35071 gttagcaaga ttgctagctt tgtgggacag atggtttcag gagggtgcga cctgattcga aacttcatta
 35141 gtggatttgg gtcaatgatt ggttcagctg tctctaaaat tggcagcatg ggaacttcaa ttgtttctaa
 35211 ggttacttga ttcgctggac aaatggtaag cgcagggggt aaccttgttc gaggatttat caatggatc
 35281 agttccatgg taagtctgc ggttaagtgc gcggtcaata tggctagcag tgcattaaat gccgttaagg
 35351 gattcttagg tattcactct ctttcacgtg tcatggagca gatgggtatc tatcgggtc aagggtctgt
 35421 aaatggtatt ggttaactga ttcgaactac acgtgacaa gctaaagaaa tggctgaaac tgttactgaa
 35491 gctctcagcg acgtgaagat ggaatttcaa gaaaatggag ttatagaaaa ggttaaatca gtttacgaaa
 35561 agatggctga ccaacttccc gaaactcttc cagctcctga ttcgaagat gttcgtaaag cagccgggtc
 35631 gcctcgagtg gacttgttca atacaggaa tgacaaccct aaccaactc atgcacaaat taaaaacaat
 35701 caaggcgagc aaacggttgt caacattgga acaatcgtag ttcgaaacaa tgacgacgct gacaaactgt
 35771 cgagaggatt gtataataga agtaaaagaa ctctatcagg gtttggtaac attgtaacac cgtaaaggag
 35841 aaatagatgg ctagcagaca gcgctattg gtcgacgga ttgacctgt cgacaaaggt gcaaccgtgc
 35911 tagaatatgt aggaactcact ttgcaggat ttaaggactc aggtatttaa aacctgaag gcataagcgg
 35981 agtatagat tctccgtcta atgctatgt cgccttactt ggaagcgtga ccttaattgt ccacggagaa
 36051 accgaaaagc aagttaatca aaatataggg agtttcaaac aatttattcg ctggaagtc ttttggagaa
 36121 ttcgacact tgaagacctt ggaatactat ccttctaggt aaatttagga gaaacggagc aaggaaaact
 36191 tgtagacgtt caagccttta aagatacttc ccttctaggt aaatttagga ttcagttcaa agatgcttac
 36261 gagttagcgg actcaactgt tcgaaaggtt tataagtttc gggaggggat agcttacctc
 36331 acccagggaag acctactcga caatttagag tagaaataag aactacttct caaatcaag gatatttctg
 36401 aartggcgaa aaaagttcag gacagtttgt tgagttcgg actaattcag tattgatgga aagtggctcg
 36471 attatattc taaatcttgg accttttgaa cttattaaaa tttagcagtg aaatcaagcg actaacttat
 36541 ttgatatac taaacgaggg gcattcttca agattctcaa tggaaattca acaattacca ttgaatccg
 36611 agccgatgac gcagcagctt ggacctctac tcttcccgct caagttgaac tgtttctaaa tccgtcttac

36681 tattagaaag ggaatatatg attgacaata atttacctat gagtccaatt cctggcgaaa ttgttcaagt
 36751 atattgaccaa aacttcaatc taattggagc aagtgatgaa atcttttagca agcattacga agacgaaatt
 36821 gtgactcgag ctcgaggaaa agaaactttc acttttgaaa gtattgaaac ctcatctatc tatcaactac
 36891 taaagggtga aaacattatc cagtatggag gaagatgggt tcgaattaaa tatgcttcagg acgtagaaga
 36961 tgcctaaagg cttaccaagt ttacctgcta cgcattatgg tatgaactag cagaaggcgt gcctaggaag
 37031 ttgaaacacg ttgcttcttc tgcaggcgct gtcgcgctag atattatcaa agacgcaggt gaatgggttc
 37101 gactagtttg tcttctctgac ggtgctaaca aacaggttcg aagcataaca gccgcagaaa attcaatgct
 37171 ttggcatctt cgatatcttg caaagcaata caatttagaa ttgacatttg gttatgaaga aattatcaag
 37241 caagagggtta gaattgttca aacggttcta tttcttcagc ctatgtcga gtctaaagta gactttcttc
 37311 ttgtagtga agagaatttg aaatatgtca ctaggcagga agattctcga aacctgtgta cggttcaaa
 37381 gttgacaggt aaaaaggaaag aaggcagtcga agagccttta acgtttgctt ctatcaacaa tggaaagtga
 37451 tatctcattg atgtrtcttg gtttactaca cgccacatga agcctcgata tattgctaaa tctaaaagcg
 37521 acgaacattt tagaattaaa gaaaatttga tgagtgtctg cgtgtcttat cttgacatct acagtcgccc
 37591 actaattgga tatgaggctt cagcgttctt ttatacaaa gttcctgact tgcatactac tcaactaatt
 37661 gtcgacgacc attatgatgt tatcgagctg cgaagatat ctgctcga aaattgactac gacgaccttt
 37731 caaactctac tatcattttc caagaccttc gaaaagactt gatggacttg cttaaatgagg acggcgaaag
 37801 agtctcttca ggggaaactg taatagagtc ccaagttggt attagatacg cagatgacat tttagggact
 37871 aattttaatg cagaactctg gaaatcacct ggtgtcttta atactaataa gaaacgagc gaattagtct
 37941 ctgacgactt tacatggatt cgactagaag gtcttaagg tgacgaggt ttaccgggag ctctggggcg
 38011 tgatggagtc gacggtgtac ctggaaagag cggagtaggg atagcagata cagctatcac ttatgctgta
 38081 tccgtttccg gaacgcaaga gcttgaatat ggtgagagc aacaagttcc tgaactcata aaaggtcgat
 38151 tcttgtggac taaaacattt tggagatata ctgacgctc acatgaaact ggatactccg ttgcctatat
 38221 agggcaagac ggaatttccg gaaaagacgg aatcgaggt aaggacggag taggtatagc cgcacacgaa
 38291 gtcctgtagt caagtctgcc atctgctact gaagctccag ctgggtgagt gtctacgcaa gttcctaccg
 38361 tccaggtggg tcaagtatta tggactcgaa caagatggcg ctacactgac caactgatg aaattggata
 38431 ttcaagtcca agaattggcg agcaggggtc taaaggtgac gcaggtcgtg acggtattgc aggaagaac
 38501 ggaacagggg tgaagtcaac ttcagtttct tatggaatta gtcccactga tctgtcgatt cctggagatg
 38571 gggcttcaaa agtctcttct ttaatacaag gtcaatatct ttggactcga actatttga cctataccga
 38641 ttcaactacc gaaacgggct atcaaaaac ctacattcca aaagacggga atgacggtaa aaatggaatt
 38711 gctggtaagg atggggtagg aatcaagctc acgaccatta cctacgaggt ctcaacctca ggaacagttg
 38781 cgctactctc aaattcgact tctgctattc caaatgttca accgggattc tcttrgtgga cgaaaactgt
 38851 ttggaactat actgatgaca ctacgaaac aggttactca gtttccaaga taggtgaaac aggtcctaga
 38921 ggagttcaag gtcttcaagg tcttcaaggg ctccaaggaa tctctggacc tgcaggagct gacggacgtt
 38991 cgcaatatac tcacctcgtt tctcttaata gtccaaacgg tgagggattt agtcatctg acagcggagc
 39061 agcatatctc ggtcagttac aagatttcaa tcccgctcat tcaaaagacc ctgcagccta tacatggagc
 39131 aaatggaaag ggaatgacgg agtcaagggt ataccggga agccagggcg agacggtaag actaatattt
 39201 tccatatagc ttacgttcca agtgcagagc gatcacgtga gttcagtttg gaagataaat atcaacata
 39271 tatgggttat tactcagatt atgagcaagc agatagcagg gatcgaaata agtatcgatg gtttgcagc
 39341 cttgccaatg ttcaagtggg aggtcgaaac gagttcctta attctttatt tgaatttggg ttaaaacctc
 39411 gctattctag ttcaaatcta atggacggac aagatcaaac gcaaggacag atatctgcta ctattgacga
 39481 acgtcaacgg ttcaaaagtg ctaactcttt acgacttgac tcaacatgga acggttaaac gcgaaccaa
 39551 aaactgacct tttctttagg aggatgacg cgattaggtta tccaaccca ggtgtctaat ttagaaggtc
 39621 gttacagttt ctgggctaa gcttcttaga acggagtgag cttagctgca cggccggggt atctagttaa
 39691 cgtatttacc gcaaccttaa ccgatcaatg gaagtcttac gatttttaaa tctttgacaa agttaattca
 39761 aattgtaccg ctgaagcaat ttccatgta ttactcaaa gttgttcagt gtggctcaat catattaaa
 39831 tcaaaacttg taatatctct actccttcta gtgaagcaga ggaagacctt aaatatcgaa ttgactcaa
 39901 agccgatcaa aagctaaata accaactagt gacggcactc acggaagagg ctcaactaca tgcacgaga
 39971 ctgaaggtta aggtcaaat ggaagcagta agtaacttag aaaaaggctta tgaaggtaga atgaaggtta
 40041 atgaagaagc tatcaaaaaa tgggaagcgc acccaactct agcggcaagt cgaattgaag ctactatcca
 40111 agaaacttgg gggctacggg aactgaagaa gttctgacg acgttatcga gctcttctaa tgaaggtcta
 40181 attatcggta agaacgacgg tagctctacc attaaggtat caagtacgg aattttctatg ttctcccgag
 40251 ggaatgaagt tatgtacct acgcaagggg tcatctacat cgataacggg atctttacc aatcattca
 40321 agtcggccga tttagaacgg aacaatactc gtttaatcca gacatgaacg tgattcggta ttaggataa
 40391 ggagaataac atgacaaaat ttatcaactc atacggccct ctctacttga accttactgt cgaacaaagt
 40461 agtcaggagc taacgaacaa ctctcgcga gttagtggc gagctactgt cgaccgcat ggagcttatc
 40531 gaacgtggac ttatggaaat attagttaacc ttccgctatg gttaaatggt tcaaggttcc atagcagtc
 40601 cccagactac gacacgtccg gcgaagaggt aacgctcgca agtggaagag tgaactgtcc tcacaaatag
 40671 gacgggacaa agacaatgtc cgtttgggct tctgttgacc ctaataacgg cgttcacgga aatatcacta
 40741 tctctactaa ttacacttta gacagtatcc caaggtctac acagatttct agttttgagg gaaatcgaaa
 40811 tctaggatct ttacatacgg ttatctttaa ccgaaaagt aactcttcta cgcatacagt ttggtaccga
 40881 gttttcggta cgcactggat agatttaggt aagaaccata ctactagctg atcctttacg cgtcactgg
 40951 acttagcaag gtacttacct aaatcaaggt ccggaacaa atctcgaacct ataacggaac
 41021 tacgcaaat ggtagttagc tctattcaaa cggatggagg ttcaacatcc ccgattcagt acgtcctact
 41091 ttttcgggca tttctttagt agacacgact tcagcgggtc gacagatttt aacagggaac aacttctctc
 41161 aaatcatgtc gaacattcaa gtaacttca acaatgcttc cggcgtctac ggaatcacta tccaagcatt
 41231 tcacgctgag ctctgaggtta aaaaaccaagc tatcaacgaa aacggcgga aattgggtat gatgaacttt
 41301 aatggctccg ctaccgtaag agcatgggtt acagacacgc gaggaaaaca atcgaacgtc caagacgtat
 41371 ctatcaatgt tatagaatac tatggacggt ctatcaattt ctccgttcaa cgtactcgtc aaaatcctgc
 41441 aattatccaa gctcttcgaa atgctaaggt cgcacctata acggtagag gtcaacagaa aacatcarg
 41511 caaattacct tctcgtggc cgcgttgaa actactaatt tcacagaaga tagaggttcg gcgtcaggga
 41581 cgttactctac tatttcccta atgactaaat cgtcccgaa cttagtctgt aactacgggc cggacaagc
 41651 ttacatagtt aaggctaaaa tccaagacag gttcacttcg actgaattta gtgtacggg agtcaacgaa
 41721 tcaagtgttc ttaactatga caaggacggt cgacttggag ttgtaaggt ttagaagaaa gggaaagcag
 41791 ggtcaattga tgcagcaggt gatataatg ctggaggtcg acaagttcaa cagtttcagc tcaactgata
 41861 taatggagca ttgaacaggg gtcaatataa cgtatgttgg aataagcgtg aaacagagtt tacatggcga
 41931 agtaacaaat acgaggacaa cctacggga actcaggtg aatggggact atttcaaaat tctggttag

42001	atagctggaa	aatgggtcaa	tcccttcatta	caatgtcagg	aagaatgttc	atcaggacag	cgaacgatgg
42071	aaacagctgg	agacctaa	agcggaaaga	gggtctat	aagcaagact	tcgaacagaa	taattggcag
42141	aaactgtgtc	ttcaaatgg	gtggaaacct	cactcaacct	atggcgacgc	attctattcg	aaaactcttg
42211	acggcatagt	atatttgaga	ggaatgtgc	ataaaggact	tatcgacaaa	gaggctacta	ttgcagtact
42281	tcctgaagga	tttagaccga	aagtttcaat	gratcttcag	gctctcaata	actcatatgg	aaatgccatt
42351	ctatgtatat	acactgacgg	aagactttgtg	gtgaaatcga	atgtagataa	ttcttggtta	aatttagaca
42421	atgtctcatt	tcgtatttaa	tttgagctga	aatcatgtta	taatatatttt	tagaaaggag	gtgagaacta
42491	tggtgaacct	tacaaaatcg	cgccaaattg	tgccagaggt	cactatttga	caaggagctg	aaaagaaact
42561	tgtaaaaaca	acgattgtga	acattgatgc	aaacgcagta	tcaaccgtct	ctgaaactct	tcactgacca
42631	gacttgtatg	ctgccaacgg	tcgagaacct	cgagctgacg	agcaaaaact	tcgcgaaact	cgctacgcga
42701	tcgaagatga	aattctagct	gaacagtcga	agactgaaac	agctctaaca	gctgaataag	gaggcgctca
42771	tctatgccaa	tggtggctaaa	cgacacagca	gtcttgacga	cgattattac	agcgtgcagc	ggagtgtcta
42841	ctgtccctact	aaataagtta	ttcgaatgga	aatcgaataa	agccaagagc	gttttagagg	atatctctac
42911	aaactcttagc	actcttaaac	agcaggtcga	cggyattgac	caaacgcagc	tagcaatcaa	tcacaaaaat
42981	gacgtcatct	aagacggaa	tagaaaaact	caacgttacc	gtctttatca	cgacttaaaa	aggggaagtga
43051	taacaggcta	tacaactctc	gacctttta	gagagctctc	tattttattc	gaaagtata	agaaactctgg
43121	cggaatgggt	gaagttgaag	ccttgtatga	aaaatacaag	aaattacca	ttaggaggga	agatttagat
43191	gaaactatct	aacgaacaat	atgacgtagc	aaagaacgtg	gtaaccgtag	tcgttccagc	agcgattgca
43261	ctaattacag	gtcttgagc	gttgatcaa	tttgacacta	ctgctatcac	aggaaccatt	cgactctctg
43331	caacttttgc	aggtactgtt	ctaggagttt	ctagccgaaa	ctaccaaaag	gaacaagaag	ctcaaaaaca
43401	tgaggtggaa	taattgggagt	cgatatgaa	aaaggcgttg	cggtgatgca	ggccccgaa	ggctcgagtat
43471	cttatagcat	ggactttcga	gacggtcctg	atagctatga	ctgctcaagt	tctatgtact	gtgctctccg
43541	ctcagccgga	gcttcaagtg	ctggatgggc	agtcaatact	gagtcacatg	acgcattggc	tattgaaaac
43611	ggttatgaac	taatttagtga	aaatgctcgg	tggtgatgta	aacgaggcga	catcttcatc	tggtggagcga
43681	aaaggtgctag	cgacggcgct	ggaggtcata	cagggtatgt	catctgacag	gataacatca	ttcactgcga
43751	ctacgcctac	gacggaattt	ccgtcaacga	ccacgatgag	cggtgggtact	atgcagggtca	acctactac
43821	tacgtctatc	gcttgactaa	cgcaaatgct	caaccggtcg	agaagaaact	tggtggcag	aaagatgcta
43891	ctgggttctg	gtacgctcga	cgaaacggaa	cttatccaaa	agatgagttc	gagtatatcg	aagaaaacaa
43961	gtcttggttc	tactttgacg	accaaggcta	catgctcgct	gagaaatgg	tgaacacac	tgatggaaat
44031	tggtattggg	tcgaccgtga	cggatatact	gctacgtcat	ggaaaacggat	tggtgagctca	tggtactact
44101	tcaatcgcg	tggttcaatg	gttaaccggt	ggattaaagta	ttacgataat	tggtattatt	gtgactgtac
44171	caacggcgac	atgaaatcga	atgcgtttat	ccgttataac	gacgggtggt	atctactatt	accggacgga
44241	cgctcggcag	ataaacctca	attcacctga	gagccggacg	ggctcattac	tgctaaagt	taaaatatag
44311	agaggaggaa	gctcttttct	taatatgtgt	tctcttaatc	ccgcaagggt	tcgacctg	gggtgtttgt
44381	gtcgatatatt	actctattta	cttattcgaa	gatttcaatt	ataattaaat	agtcaacatg	attcatgatt
44451	gttgatatga	ccctttccgc	cctacataat	ttgtggggcg	tttatttttt	ataaaaaatt	tttacaanaat
44521	gcttgacaac	attcactcat	tatcgataaa	tacaattata	aaaataaata	aaagccgaa	gcgaggaggga
44591	cattatgtca	aaaattaaat	tcgaaaacct	taaaaaaggc	gatgttgtgc	tacgagctaa	atctcaaacg
44661	aagtttaaaa	tcggtttcaat	tttagcagac	gaaaagaaag	cagaccttga	atcattagaa	gacggagggtg
44731	aaacttcacct	ttcagcttca	actctcgaa	gttggtacac	aatgggaagt	gaaactgaac	ctaaaaaaga
44801	agaagctgct	aaacctgcta	aaaaggctgc	tcctgcagtt	gctcgacctg	ctcgaaaagg	tagagctcgt
44871	cccaaaccta	aaaagaaagt	ccttgaggaa	gaaattctctg	aaagttaagg	acagccggaa	gaagtgtggt
44941	cagtttagtg	gaaatctact	gttcgaaaac	ctgctcctaa	aaaagaaagc	gtgatggcga	ttactaaagg
45011	tcctgaaagt	cgaattgttg	aagcctttcc	tgctctact	cgaatcgta	ctcagctcta	catcgctact
45081	cgctctaaga	agaactctgt	tactatcgaa	gaaactcgaa	aaagtgtttc	tattggagtt	cgcgcaaaa
45151	ggttgacaga	agaccaaaag	aaacttcttg	catctattgc	tcctgcatct	tacgaatggg	acttgacgg
45221	aatttttaa	ctcgtcaagg	aagaagatat	tgacaccgca	atggaaatga	ttgaagcttc	tcacctttct
45291	tcgctatgat	tgaaatcggt	atagcacgtt	cgaaagctag	gcgaggtcga	accctattta	ttgaaacatg
45361	ggcaagcact	gatgaagatg	cagttaaaat	ggcagaaaag	atttccagct	tgcccaatgt	agtcgagacg
45431	ctcttcaata	acttcgaact	accttataag	tatttcaata	atgttataga	cgctctagat	gaatgggagc
45501	ttcacatctt	cgccgaactt	gataaagatg	ttcaagacta	cattgactct	cgaaaccgaa	tagcttcttc
45571	aagcaatgag	cagttttcgt	tcaagactac	tccattcgcg	caccaggttg	aatgtttcga	atacgcacaa
45641	gagcatccat	gtttcctttt	aggcgatgag	caaggtttag	ggaaaactaa	acaggcaatt	gatattgcag
45711	ttagcaggaa	ggcaagtttc	aaacattgtt	taactcgatg	ttgcataatca	gggctcaaat	ggaattgggc
45781	aaaagaagta	ggtattcatt	caaattgagtc	agctcatatt	ttaggaagtc	gagtcactaa	agatgggaaa
45851	ttagtattg	acggagtttc	taaacgggca	gaagacttgc	ttggtggcca	cgacgaattc	ttccttatca
45921	ctaacttga	aactcttcgc	gatgctgtgt	tcattaaata	cttaaatgaa	ctgacaaaaa	gcggagaaat
45991	tggaattggt	attattgacg	agattcacia	gtgtaagaac	ccttcaagta	agcaaggggc	ttcaattcaa
46061	aagctccaaa	gttattacaa	gatgggactt	acaggaaact	ctctaagtga	taacccaatc	gatgtattca
46131	atgttatgaa	gtggctaggg	gcggacatc	atacactgac	tcagttcaaa	gagcgatact	gtatcgtcga
46201	ccagttcaat	caaactcactg	gatattcgaa	tctagctgaa	cttcgcgagc	ttgtcaacga	ctacatgctt
46271	agaagaacga	aggaagaagt	tttagacctg	cctgaaaaga	ttcgagtcac	agagtatgtc	gacatgaaat
46341	cgaacacgtc	aaaaatctat	aagggaagttt	tgactaaact	tggttcaagaa	atagataaag	tcaagctcat
46411	gcctaaacct	ctagccgaaa	cgattcgact	tcgacaagcg	actggaaatc	cttcgatttt	aactactcaa
46481	gatgtcaagt	cttgcaagtt	cgaaagatgt	atcgaaattg	tcgaggaatg	tatccagcaa	ggaaagtctt
46551	gcgtgatatt	tagcaattgg	gaaaagggtta	ttgaacctct	tgctaaagata	cttcgaaaga	cagtcacaa
46621	caacctgggt	acaggagaaa	ccgcagataa	gttcaacgaa	attgaagaat	ttatgaatca	cagaaaggct
46691	tctgttattt	taggaactat	aggtgcgcta	ggaaacaggt	ttactttgac	gaaagcggat	acgggtattt
46761	tcttagatag	tcctgggaca	cgccagaaa	aggaccaagc	cgaagatagg	tgctatagaa	tcgtcgaatg
46831	aagttctgtc	actatctaca	cgctgtgcgc	caaaggtaact	gttgacgaac	gtatagaaga	ccttattgaa
46901	cggaaaggag	aattagcaga	ttatatcgta	gatggtaagc	ctatgaaatc	taaaattggt	aaccttttgc
46971	atatcctgct	taaatagaat	gaaaactact	tccatattaa	ggaaagacac	taaaaggaa	ccggacagga
47041	acggaagaaa	aactgcactc	gaactagctc	aagagattga	tatgtcacct	agtgaagtag	cagagctcct
47111	tcaaatctct	gaaaggacgg	caaccagaat	tttaaaactc	gacaaaactgc	tcaacaaaga	gcaatgctca
47181	ataatagaaa	ggtatataaa	tgaattcac	tgaaaggaaa	aattgggata	aagtgggaga	gatattgcaa
47251	atgttgaaac	gctctctatc	tacgattaat	gtttggtagt	aagcaaaaaga	cttcgctgaa	gaaaaataca

47321 ttcacttccc gtttgttctt cctgaacctt gaacagacct tgaccatcgt ggttctcgtat tctgggatga
47391 cgaaggcgtg aacaaactca aacgactttag aacagacctt atgcgcggtg acctggcatt ctacactcga
47461 actcttctag ggaataactga aaggggaagca attcaagaag atgctaaagc atttaaacgt gaacatggat
47531 tggagaatta aatgaaattt gaagatgaaa aacagttcat cgtctgaact gaagaagccg gtgaattaaa
47601 tgcaccaaa ggcgacatgg agaaacaagt caaaagtctt cgtgatgctc taaaagagta catgaaagaa
47671 aatgacattg aatctgctca aggttaagcac ttttctgcta cctctacac gacagagcgc tcaactatgg
47741 acgaagaacg cttgaaagaa attatcgaaa aattagttga cgaagccgag acggaagaaa tegtgtgaaa
47811 actttcaggg cttatcgaa acgaacctgt catcaatag aaactctctg aggatgatg tttatcacggc
47881 gagattgacc aagaagcaat tcttccagca gttgtcattt ctgttacaga aggcattcgt tttggaaagg
47951 ctaaaattta gcgataattt tgggtctcgc acgtttttag ggttagcaga atccaatcac accacttcgc
48021 caggcaaccg ctgtctcgct caattttaga aggttaatat tataccataa ggaggagata agtggcaagg
48091 caaagaatag gcaattcagg aaagcctaaa aatgaaattg aactaacatt caaagacaag cctaaaactc
48161 gttctacctt attcaagaag gacgtggcaa caggtctttc aaaagtcgag catgattatt tccaattagt
48231 tgaagcaact aacggaaaac aattcgaaac taatatgaag caggtgctat ctttctttat agtcacaggc
48301 gaattttatt tcaatattaa gtgcactgat tataactggt tcaacttttc gacgactatg aaaaatgttc
48371 gaactttatt aaactattgag tcgaacattg aactttgtcg atttttagct gaaagttttg ttaaatatga
48441 aattgttcga aaaagattga acctaaagca aaggttcata acggtctcga ctttcaaaag agcctggatt
48511 ttggagcaac tcgaaggaaa aacgggttca aatttcgaag gattttatta gtttagtga ctacttttag
48581 attttttaa atgtgtttta caaaatgacc tcaataggcg tataatttat caactctgat tctttcgggc
48651 cgttatatat acaccaataa tcgagaataa ataattata gtatcgaaaa tataaaaagg agaaaagtgt
48721 gaaaatttga ctgatagaat atggaagaaa aagttaaatg accttttcga gagaagtggg ctacctcaaa
48791 agtatttcga acctcaagt ttagtcgaac gaaaagccga caaggaatgt tgggaatggc tagaagctgt
48861 tcgagcaaat atagtcgaag aagttcgaaa cgggtcttagc attgttatgt cttcgaatag tgcgggaat
48931 tggaaaaacta gctggggcgt tcgacttttg caacgctatt tagcagaac tgcacttgac ggaagaattg
49001 ttgagaagg aatgttttga gtgtcagctc aactattgac tgagtctggc gactataatt attttcaaac
49071 catgcaagaa tttctcgaa gtttctgagc ccttaagact tgtgagctat tagtcataga cgaaataggc
49141 ggaaggttctt taaccaagc ctcttatcct tatctgtatg acttggttaa ttatagggtt gacaataact
49211 tgtcgactat ttatcgact aattatctg acgatgaat tattgacct ttaggcaaaa ggtttatag
49281 tcttatatat gatacttcag tgggtctaga ttttcaggca agcaatgtaa gaggattgga ggaagcgaa
49351 attgaatcat agatatagta acatcacac tatttttctt tggcagattg tctttcttctg tatttctgctc
49421 gcggtgtcct attgtgcagg agtcataat gagcgagagt ctcaagataa ggtgattcaa agttataagc
49491 agaaaagaaa gtcagccgtc tacttgacag tcgatatgtc aggaacttgg ctggaagtg ctcggggagc
49561 caaggaaggt cctctctaca atgaaaaggg acagcatgta ggaataatga aagaggtggg agagtgatac
49631 agcttcaagt cttaaaataa gttctcgaag aaaagagctt atccatttta gaaaataatg gaattgacca
49701 agaatacttc acggattatt tagacagata tcaatttatt caagaacact tttcgagata tggaaagatt
49771 ccggagcagc aaactattct cgaccatttt cctggattcg aatttttcga aattggcgaa actgatgaat
49841 accttatcga caagctaaaa gaggagcatc tatataattc acttgttcca attttaacgg aagcggctga
49911 ggacattcaa gtagatagta acattgcgat tgcgaatata attccaaaac tagaagaact tttcaactcg
49981 tctaaattcg tagggcgact agacattgct cgaatagcta aacttcgact agactggggc aatactatta
50051 gaaacatcga cggtaaaaga cttggaatat cgacagggtt tgaactattg gacgactgc tttggaggctt
50121 acttctcgtt gaggatttga ttgtcataat ggctcgacct ggacaaggta agtcgtggc tattgataaa
50191 atgcttgcga ctgcttggaa gaacgggcat gatgtccttc tatatagcgg ggaatgagat gaaatgcaag
50261 ttggtctcgt tatagatact attctttcga atgtttagcat caattcaatt accaaaggga tttggaacga
50331 ccatcagttc gaaaaatag agggaccat atcaagcaatg actgaggctg aaaattccct tgtggtagtc
50401 acgcccctta tgattggagg aaagaacctt acccctgcaa ttttagatag catgatattc aaatatagac
50471 catctgtggt ccaaatgctc agcagagtta tcgatgagca gtccttacc agcaggagc agaaagcaat
50541 ccagtagcgc aacatcacca tggacctata taagatttct gctaaatatg gaattcctat tgtgcttaat
50611 gtccaagcag ggcgttcggc taaaactgaa ggcgtgaaa gtagtgaact agaactata gcagaagtg
50681 atggagtagg tcaaatgctc agcagagtta tcgctatgaa gcgtgacgaa aaatccggca tacttgaaact
50751 atctgtcgtt aaaaaccgat atggcgaaga ccgaaaaatc atcgaatata tgtgggacgt tgaactgga
50821 acctatactc ttataggatt caaagaggaa ggcgaagaag gaactgaaa aggcgaagc tctccattga
50891 aagcaaaagc ctctaggtcg actgctcgtc ttcgaagtaa ggttacagg gaaggagttg aagcattttg
50961 atgaaagtaa atggtcttca aattgaagcg actcctgaac aaataattga aaaactttcg agacaacttg
51031 aagacgaagg aacattcatt tttagacgaa ctaagtcgct tggaaagcaac tatcaattct catgcccgtt
51101 tcatgcagga gggactgaaa agcatccctc ttgtggcatg agtagaaatc cttcttattc aggaagtaag
51171 gtgacggaag ctggaacggt tcactgtttc acttgcggct acacttcagg actaaactgaa tctgctcga
51241 atgtattagg tcgaaacgat ggagggttct atggaaaacca gtggctgaaa aggaattttg gaacatctag
51311 cgaagtagtt aggcaaggcg tcagccctga agcgtttcga agaaatggga gaactgaaa agtcgagac
51381 aaaaactatt ctgaagagga acttgataaa taccggttta ttcactctta tatgtatgaa cggaaattga
51451 cggacgagct catcgagatg tttgatgtag gttatgaaa actgcatgat tgcacacct ttcagtagc
51521 gaacctcaag ggcgaacagc tattcttcaa ccgtcgaagt gttcgttcta agttttacca gtacgggtgaa
51591 gatgacctta aaacgggaat tctttatggc caatatgagc ttgtagcatt tcgagactat tttgaaaac
51661 ctattagtcga agtatctcgt actgagtcgt ttatcaactg cttgactctt tggtaatga agattccage
51731 agtcgctctt atgggagtag gtggaggaaa tcaaatcaat ttactaaaac gacttcttta tagaaatatt
51801 gttctagcac ttgacctga taacgctggg cagacagcgc aggaataact ctaccgacag ttaagcgaa
51871 gcaaggtcgt tagatttttg aactacccta aagagttcta tgataataag tgggataata accagcatcc
51941 ggaattatta aattttaatg atttagtctt gtgaaaattc atttattatc gtataataaa gtttagaaat
52011 tttaaaaaga ggtcatatca atatgaaaga agcgaataga ctagtcttca gctatgtagg attcgaatgc
52081 tggactgacg aagaatgtat caggaaacttt gaactagacc ctgatctgtc aattgctgtc gcttaccatc
52151 gttatttttg gatgctttat tctatgcaa aaaggtttta atgcttatct cgacatgaca ttgaaagcat
52221 tgcattcgag actatttcaa aatgtttggc aacgttcaaa tcaaaccaag gggccaaagt ttaactttac
52291 cttacaagac tcttcaagaa tagaatagtc ttgaaatata ggtacctaaa tgcacttcc atgaatcgaa
52361 attggtatgt agaatgacg ttcgatagcg tttcgacaaa tgaagaaggc gacgatttta gtatcctatc
52431 gacagtttgc tattgtgaag actacggaaa aattgaaatt gaagcaagtc ttgacttcat gacgtttctc
52501 aatacagagt atgcttatat ctgctcgtc attcaaaacg gtccttcagt aagcgacgca gaaattcgcc
52571 gtgaaatttg agtaagcagg tctgctatta gtcagcttaa gaagtcacta aaaaataaat taaaagattt

52641 tatataactg gtttacaat cactgaatt tctgtatcat tatatatgaa aggacaaact ttgaaacctt
 52711 aaaaacttca aaaatcttcc aaccattaaa aacttataaa ggagaatcga tatgggaaaa gtatcaattc
 52781 aaaaatcagg aacatttagc tcagggtcta ataacgagtt ttccacactc gctgaccacg gtgacagcgc
 52851 aattgtcact ctattgtatg atgaccgcga aggcgaagac atggattatt tctgagtcca cgaagcagac
 52921 gttgacggtc gtcgacgcta tatcaattgc aatgctattg gcgaagacgg ggaacagatc catcctgata
 52991 actgtccatt atgccaatc ggattccctc gtattgaaaa actatttctt caactttaca accatgatac
 53061 gggaaaagtt gaaacatggg accgagggcg ttcttatgtt caaaagattg ttacatttat caataaatat
 53131 ggaagccctg tgaactcagc ttttgaaatt attcgttcag gaggtaaggg tgaccaacga actacttatg
 53201 aattccttcc agagcgtccg gaagacagtg ctactcttga agattttcca gaaaagagcg aacttcttgg
 53271 aactctaatt tttagacctc acgaagacca aatgtttgac gtggttgacg gcaagttcac tcttcaagaa
 53341 gagcgttctt caagtctgtc aaattcacgt agaggagcat ctcttcgccc tagacgaggt tccggtcgag
 53411 aatcttcaca aggtcgaaca gctgaaagaa ctcttccagt tagtcgaaga actctccaa cacgaggtcg
 53481 aggatctctaa catgagggcg cgagccctct ttattattga ttaagaaagg gaaaataatg gcacaaaaag
 53551 gactcttctg tgcgaagcct cgttctagca agaagaacga tgcctcagtt cttgctcaac ggaataacag
 53621 gaagcccgca gttgaggtta ctacatttcc ataaaggagc cttaaaggagc agttgtctag agctctgact
 53691 ctttcaacta ggaattcttg acacgttctt gatagacttg agttaatcac tgaggaagca aaactcgagc
 53761 agtatgtaga caaatgatt gaagacggaa taggttctat tgacgtagaa actgatggac tgcatactat
 53831 tcaagtgag ctggcaggag tctgcttcta ctcacctagt caaaaaggaa tctatgctcc tgtcaatcat
 53901 gttagcaata tgacgaagat gcgaattaa aatcaaat tctctgagtt catgagaaa atgcttcaac
 53971 ggattgtaga ttcagggaat cctgtcatct atcataatc gaaatttgac atgaaatcga tttattggcg
 54041 actcggcgct aaaaatgaat agccagcggt ggatacatat tttagccgaa tgccttttaa tgaatacag
 54111 tctcaccagt tgaagaatg tcaactctaa tatgttagga acgaagaaaa cgcagaggtt gcaaaattta
 54181 atgacttatt taaaggaaat ccttttagtt taactctccc tgatgttgcc tatatgtag cggtctatga
 54251 ceetttgcaa actttcgaac tctatgaatt tcaagaacaa tacttgactc caggaaactga acatgtgaa
 54321 gaatataaac tggaaaaagt ctcattgggtt cttcataata ttgagatgcc tctaataaaa gttctcttcc
 54391 acatgggaagt ctacggtgct gacttagacc aagataagct ggcaagaatt agagaacagt ttactgccaa
 54461 tatgaacgag gctgagcaag agtttcaaca gcttctcagc gaatggcagc ctgaaattga agaacttcga
 54531 caaactaatt tccagagcta tcaaaaactc gaaatggatg caagaggtcg agtgacgga agcatttcca
 54601 gctctactca actagcaatt ctgttttatg atatcatggg attgaaaagt cctgaaaagg ataaacctag
 54671 aggaacaggg gaaagtattg ttagagcatt tgaataacgat atctcaaaag cacttttgaa atatagaaaa
 54741 tatgcaaaat tagtttcgac ctatacaaca cttgaccaac accttgcaaa gcttgacaat cgaattcaca
 54811 ctacattcaa acagtacgga gctaagacag ggcgtatgtc aagtgaagat cctaacttac agaattatcc
 54881 ttctcgggtg gagggtgtag tagttcgaca aatctttgca gccagtgaag ggcattacat tattggtagt
 54951 gactactctc aacaagaacc tctgttcttg gcggaattaa gtggcgacga aagtatcgca catgcttacy
 55021 aacaaaacct ggacctatat tcagtattcg gttcgaaact ttatggtgtt cctatgaag agtgtttaga
 55091 gttctatccc gacggaaacga ctaacaagga aggaataact cgaagaatt ctgtcaagtc cgttctttta
 55161 ggtcttatgt acggccgagg ggctaactca atcgtctgagc agatgaatgt atctgtcaa gaagcgaata
 55231 aggttattga agatttcttc accgagttcc ctaaaagtggc agactatctc atattcgttc aacagcaggc
 55301 gcaggacttg ggaatatttc aaacagctac cggctgaaga agaagggctc ctgatatgag tcttcttgaa
 55371 tacgagttcg agtatatcga cgttagcaag aacgaagatt tgcacccctt taactttgac gcagaccaac
 55441 agatggacga tactgttctc gaacatatta tgcgaataa ttggggccag ctgatagag cctggggagt
 55511 taagaagaag caagaataa aagaccaggc aaaagccgaa ggaattctta ttaaggataa cggaggcaag
 55581 atagctgatg ctacagcgca atgtttgaac tcagttattc aagggaatggc agccgacatg actaagtacg
 55651 caatgattaa ggtacacaat gacgctgaat tgaagaatt aggatccat ttaattgattc cagttcacga
 55721 tgagtacta ggtgaggttc ctatcaagaa cgcgaacgg ggagcagaaa ggttgacaga agttatgact
 55791 gaagcagcca aggacattat tagtcttcca atgaaatgtg accccagtat agtagaaga tggatgggtg
 55861 aagaatttga aatctaaaat ctactcagtt gcataataa ttctagtagt tattgccaac cttgtgacaa
 55931 tttatttcca acctttaaat gtgaaaggaa ttttaattcc tccaagcagt tggtttatgg gattcacttt
 56001 cctgcttata aatctaataa gcaagtacga gaagccaaaa tttagcaggtt ctttgatag ggtagggtta
 56071 ttcttacct cgttgatttg ctttatgcaa aacctaccac aatcgtttgt cgtggcttca ggaagtgcac
 56141 tttggataag tcaaaaagca agtgtcttta tttcgacaa gctctgaat aaattagact cgaagatttc
 56211 aaatgctttg tctagcaaca tgggttctat tatagacgca accatattgga tttcattagg actgagctct
 56281 cttggaattg gaacgggttc atatatatag attcctgtag ccgtactagg ccaagttcta gttcagttta
 56351 tcttgcagtc aattgcttcg agatatttga aaaagttagt aggaataatc ctgattatct tgcagtcatt
 56421 tgcctcagaa tatttgaaaa agtagtcagg aaaattctct attatttttt tacaataaac gcttgacttt
 56491 attcattcat tattat

Table 29

Phage dp1 ORFs list

nb	Name	Frame	Position	Size (a.a.)	Key words
1	dp1ORF001	2	36698..40390	1230	Putative tail;
2	dp1ORF002	1	32386..35835	1149	Tail;
3	dp1ORF003	3	53538..55877	779	DNA polymerase I;
4	dp1ORF004	3	40401..42440	679	Minor structural;
5	dp1ORF005	1	23674..25434	586	
6	dp1ORF006	2	45296..46987	563	SWI/SNF Helicase;
7	dp1ORF007	3	22230..23621	463	Terminase;
8	dp1ORF008	1	49824..50961	445	DNA Helicase;
9	dp1ORF009	2	13160..14404	414	
10	dp1ORF010	2	8699..9859	386	RecA;
11	dp1ORF011	3	28017..29096	359	Major head;
12	dp1ORF012	3	5346..6419	357	DNA pol. III beta;
13	dp1ORF013	3	10215..11240	341	DNA pol. III gamma and tau;
14	dp1ORF014	3	50961..51974	337	DNA primase;
15	dp1ORF015	1	3793..4728	311	
16	dp1ORF016	3	43413..44303	296	Amidase;
17	dp1ORF017	1	11242..12081	279	
18	dp1ORF018	3	35847..36686	279	
19	dp1ORF019	2	12161..12967	268	
20	dp1ORF020	1	1864..2658	264	exsD; Coenzyme PQQ;
21	dp1ORF021	2	2504..3295	263	GTP cyclohydrolase;
22	dp1ORF022	2	30896..31675	259	
23	dp1ORF023	2	6419..7195	258	
24	dp1ORF025	-1	18026..18778	250	
25	dp1ORF024	3	25992..26738	248	
26	dp1ORF026	2	21512..22252	246	
27	dp1ORF027	1	52762..53490	242	
28	dp1ORF028	3	44595..45299	234	
29	dp1ORF029	2	662..1348	228	exsB;
30	dp1ORF031	3	26943..27611	222	
31	dp1ORF030	-2	19423..20088	221	
32	dp1ORF032	1	52033..52647	204	
33	dp1ORF033	2	7670..8239	189	
34	dp1ORF035	-1	16859..17425	188	
35	dp1ORF036	1	48808..49362	184	DNA replication;
36	dp1ORF037	1	55855..56388	177	
37	dp1ORF034	2	131..652	173	
38	dp1ORF038	3	1350..1871	173	exsC; 6-pyruvoyltetrahydropterin;
39	dp1ORF039	3	3306..3803	165	Citulline biosynthesis;
40	dp1ORF040	1	7192..7683	163	
41	dp1ORF041	3	8208..8699	163	dUTPase;
42	dp1ORF042	1	48082..48561	159	
43	dp1ORF043	1	31699..32154	151	
44	dp1ORF044	-1	25211..25666	151	
45	dp1ORF045	2	25340..25777	145	
46	dp1ORF046	3	42774..43202	142	
47	dp1ORF047	1	47542..47961	139	
48	dp1ORF048	-3	16308..16709	133	
49	dp1ORF049	-3	43620..44018	132	
50	dp1ORF050	3	15081..15476	131	
51	dp1ORF051	2	29765..30154	129	
52	dp1ORF053	-3	49917..50300	127	
53	dp1ORF052	3	30516..30893	125	
54	dp1ORF054	2	14423..14800	125	
55	dp1ORF055	3	27627..28004	125	
56	dp1ORF056	-3	18780..19151	123	
57	dp1ORF057	1	9859..10218	119	
58	dp1ORF058	3	15633..15989	118	
59	dp1ORF059	1	30154..30507	117	
60	dp1ORF060	-2	37717..38070	117	
61	dp1ORF062	-3	44940..45284	114	
62	dp1ORF063	1	47200..47541	113	
63	dp1ORF064	2	29108..29449	113	

64	dp1ORF066	-3	28566..28898	110	
65	dp1ORF067	-1	44735..45061	108	
66	dp1ORF068	3	29451..29768	105	
67	dp1ORF069	-3	20094..20411	105	
68	dp1ORF061	-3	19161..19475	104	
69	dp1ORF070	1	15973..16284	103	
70	dp1ORF071	3	38904..39209	101	
71	dp1ORF072	-2	50749..51045	98	
72	dp1ORF073	3	14262..14555	97	
73	dp1ORF074	3	32298..32591	97	
74	dp1ORF075	-1	22154..22447	97	
75	dp1ORF076	-1	5435..5728	97	
76	dp1ORF077	1	14800..15084	94	
77	dp1ORF079	-3	35007..35288	93	
78	dp1ORF081	-3	55188..55466	92	
79	dp1ORF103	2	49352..49627	91	
80	dp1ORF080	1	42490..42759	89	
81	dp1ORF082	1	44728..44994	88	
82	dp1ORF083	-1	35720..35974	84	
83	dp1ORF065	-3	51246..51497	83	
84	dp1ORF085	-3	10602..10847	81	
85	dp1ORF087	-2	29794..30036	80	
86	dp1ORF088	3	5040..5279	79	
87	dp1ORF089	-2	12256..12495	79	
88	dp1ORF273	3	56256..56486	76	
89	dp1ORF078	-3	17280..17507	75	
90	dp1ORF090	1	27037..27261	74	
91	dp1ORF091	1	43189..43413	74	Holln;
92	dp1ORF092	3	46989..47213	74	
93	dp1ORF093	-2	45538..45756	72	
94	dp1ORF095	3	8877..9089	70	
95	dp1ORF096	-1	46469..46681	70	
96	dp1ORF097	-1	38868..39100	70	
97	dp1ORF098	1	43827..43836	69	
98	dp1ORF099	3	38298..38507	69	
99	dp1ORF100	1	1597..1803	68	
100	dp1ORF101	2	19220..19426	68	
101	dp1ORF094	1	8281..8484	67	
102	dp1ORF102	2	4034..4237	67	
103	dp1ORF104	-1	21224..21427	67	
104	dp1ORF105	-2	1828..2028	66	
105	dp1ORF106	-3	10329..10529	66	
106	dp1ORF108	-1	49250..49447	65	
107	dp1ORF109	-2	31435..31632	65	
108	dp1ORF110	1	16444..16638	64	
109	dp1ORF111	1	28657..28851	64	
110	dp1ORF113	-2	17521..17715	64	
111	dp1ORF084	1	15445..15636	63	
112	dp1ORF114	2	52952..53143	63	
113	dp1ORF115	-3	5151..5342	63	
114	dp1ORF116	-1	20474..20662	62	
115	dp1ORF117	-3	24492..24680	62	
116	dp1ORF118	2	15023..15208	61	
117	dp1ORF119	2	41054..41239	61	
118	dp1ORF120	1	28387..28569	60	
119	dp1ORF121	3	39222..39404	60	
120	dp1ORF122	-1	40220..40402	60	
121	dp1ORF123	-2	21145..21327	60	
122	dp1ORF124	-3	17712..17891	59	
123	dp1ORF125	-3	49740..49916	58	
124	dp1ORF126	-3	15960..16136	58	
125	dp1ORF127	-3	13335..13511	58	
128	dp1ORF128	1	4852..5025	57	
127	dp1ORF129	2	25133..25306	57	
128	dp1ORF130	-1	16619..16789	56	
129	dp1ORF131	1	43846..44013	55	
130	dp1ORF132	-1	15137..15304	55	
131	dp1ORF133	-2	7900..8061	53	
132	dp1ORF135	3	780..938	52	
133	dp1ORF136	-1	55094..55252	52	
134	dp1ORF137	-2	36988..37146	52	

135	dp1ORF138	-3	30504..30662	52	
136	dp1ORF139	-3	11934..12092	52	
137	dp1ORF140	3	20562..20717	51	
138	dp1ORF141	-1	42767..42922	51	
139	dp1ORF142	-3	31743..31898	51	
140	dp1ORF143	-3	7410..7565	51	
141	dp1ORF144	1	36517..36669	50	
142	dp1ORF145	1	42067..42219	50	
143	dp1ORF146	1	51484..51636	50	
144	dp1ORF147	1	55207..55359	50	
145	dp1ORF148	-1	28484..28636	50	
146	dp1ORF150	-3	15033..15185	50	
147	dp1ORF134	-2	349..498	49	
148	dp1ORF151	1	28027..28176	49	
149	dp1ORF152	1	42235..42384	49	
150	dp1ORF153	2	22307..22456	49	
151	dp1ORF086	2	52760..52906	48	
152	dp1ORF154	2	18446..18592	48	
153	dp1ORF155	3	13512..13658	48	
154	dp1ORF156	3	18777..18923	48	
155	dp1ORF157	-2	13135..13281	48	
156	dp1ORF158	-3	40581..40727	48	
157	dp1ORF159	-3	30225..30371	48	
158	dp1ORF149	-3	26331..26474	47	
159	dp1ORF160	2	41324..41467	47	
160	dp1ORF181	2	52175..52318	47	
161	dp1ORF162	3	13020..13163	47	
162	dp1ORF163	3	40224..40367	47	
163	dp1ORF164	-2	6553..6696	47	
164	dp1ORF185	-3	50361..50504	47	
165	dp1ORF166	-3	23376..23519	47	
166	dp1ORF167	3	1008..1148	46	
167	dp1ORF168	-2	54205..54345	46	
168	dp1ORF169	-2	45814..45954	46	
169	dp1ORF170	-2	27460..27600	46	
170	dp1ORF171	-3	47538..47678	46	
171	dp1ORF172	-1	10325..10462	45	
172	dp1ORF173	-2	32023..32160	45	
173	dp1ORF174	-2	29629..29766	45	
174	dp1ORF175	-2	15511..15648	45	
175	dp1ORF176	-3	42894..43031	45	
176	dp1ORF177	-3	19800..19937	45	
177	dp1ORF178	-3	11787..11924	45	
178	dp1ORF112	2	32207..32341	44	
179	dp1ORF179	3	56058..56192	44	
180	dp1ORF180	-1	41042..41176	44	
181	dp1ORF181	-1	12992..13126	44	
182	dp1ORF182	-2	45235..45369	44	
183	dp1ORF183	-2	13762..13896	44	
184	dp1ORF184	-3	53196..53330	44	
185	dp1ORF185	1	22522..22653	43	
186	dp1ORF186	2	21272..21403	43	
187	dp1ORF187	2	34415..34546	43	
188	dp1ORF188	2	35609..35740	43	
189	dp1ORF189	2	42587..42718	43	
190	dp1ORF190	3	39786..39917	43	
191	dp1ORF191	-1	40865..40996	43	
192	dp1ORF192	-1	2789..2920	43	
193	dp1ORF193	-2	42325..42456	43	
194	dp1ORF194	-2	40153..40284	43	
195	dp1ORF195	-3	42453..42584	43	
196	dp1ORF196	-3	11142..11273	43	
197	dp1ORF107	1	10750..10878	42	
198	dp1ORF197	2	7484..7612	42	
199	dp1ORF198	2	24119..24247	42	
200	dp1ORF199	-1	15814..15742	42	
201	dp1ORF200	-3	47715..47843	42	
202	dp1ORF201	1	38569..38694	41	
203	dp1ORF202	2	44483..44608	41	
204	dp1ORF203	-3	22656..22781	41	
205	dp1ORF204	1	1471..1593	40	

206	dp1ORF205	1	8524..8646	40
207	dp1ORF206	1	19855..19977	40
208	dp1ORF207	1	27502..27624	40
209	dp1ORF208	2	47279..47401	40
210	dp1ORF209	3	29784..29906	40
211	dp1ORF210	-1	52955..53077	40
212	dp1ORF211	-1	20837..20959	40
213	dp1ORF212	-2	52861..52983	40
214	dp1ORF213	-2	30169..30291	40
215	dp1ORF214	-2	24151..24273	40
216	dp1ORF215	-3	35700..35822	40
217	dp1ORF216	-3	32727..32849	40
218	dp1ORF217	1	23443..23562	39
219	dp1ORF218	3	22029..22148	39
220	dp1ORF219	-1	51269..51388	39
221	dp1ORF220	-1	6215..6334	39
222	dp1ORF221	1	43507..43623	38
223	dp1ORF222	3	13212..13328	38
224	dp1ORF223	3	14055..14171	38
225	dp1ORF224	-1	13505..13621	38
226	dp1ORF225	-2	32875..32991	38
227	dp1ORF226	-2	25075..25191	38
228	dp1ORF227	-2	22999..23115	38
229	dp1ORF228	1	10450..10563	37
230	dp1ORF229	1	27634..27747	37
231	dp1ORF230	2	50723..50836	37
232	dp1ORF231	-2	30958..31071	37
233	dp1ORF232	-2	29272..29385	37
234	dp1ORF233	-3	52779..52892	37
235	dp1ORF234	1	36253..36363	36
236	dp1ORF235	2	32768..32878	36
237	dp1ORF236	-1	37418..37528	36
238	dp1ORF237	-1	1568..1678	36
239	dp1ORF238	-3	1191..1301	36
240	dp1ORF239	1	26521..26628	35
241	dp1ORF240	1	41893..42000	35
242	dp1ORF241	-1	46913..47020	35
243	dp1ORF242	-1	41231..41338	35
244	dp1ORF243	-2	51199..51306	35
245	dp1ORF244	-3	26978..27083	35
246	dp1ORF245	-3	6171..6278	35
247	dp1ORF246	-3	2724..2831	35
248	dp1ORF247	1	29641..29745	34
249	dp1ORF248	1	53560..53664	34
250	dp1ORF249	2	2012..2116	34
251	dp1ORF250	2	23837..23941	34
252	dp1ORF251	-1	39101..39205	34
253	dp1ORF252	-2	54667..54771	34
254	dp1ORF253	-3	56151..56255	34
255	dp1ORF254	-3	48375..48479	34
256	dp1ORF255	-3	9468..9572	34
257	dp1ORF256	1	15289..15390	33
258	dp1ORF257	1	28216..28317	33
259	dp1ORF258	1	44023..44124	33
260	dp1ORF259	2	4298..4399	33
261	dp1ORF260	2	24746..24847	33
262	dp1ORF261	3	288..389	33
263	dp1ORF262	3	9408..9509	33
264	dp1ORF263	-1	26951..27052	33
265	dp1ORF264	-1	6038..6139	33
266	dp1ORF265	-1	4700..4801	33
267	dp1ORF266	-2	50119..50220	33
268	dp1ORF267	-2	47266..47367	33
269	dp1ORF268	-2	12520..12621	33
270	dp1ORF269	-3	53733..53834	33
271	dp1ORF270	-3	50691..50792	33
272	dp1ORF271	-3	19638..19739	33
273	dp1ORF272	-3	1455..1556	33

Table 30

Predicted Dp-1 amino acid sequences

dp10R7001

```

36698 atgattgacaataatttacctatgagtcgaattcctggcgaaattgttcaagtatatgacaaaacttcaatttgaggaga
1 M I D N N L P M S P I P G E I V Q V Y D Q N F N L I G A
36782 agtgatgaaatctttagcaagcattacgaagacgaaattgtgactcgagctcgaggaagaaactttcacttttgaagtatt
29 S D E I F S K H Y E D E I V T R A R G K E T F T F E S I
36866 gaaacctcatctatctatcaacacttaagggtgaaacattatccagtagggaggaagatgggttcgaattaaatagctcag
57 E T S S I Y Q H L K V E N I I Q Y G G R W F R I K Y A Q
36950 gacgtagaagatgtcaaagggcttaccaggtttacctgctacgcattatggtagaactagcagaaggcttgcttaggaagttg
85 D V E D V K G L T K F T C Y A L W Y E L A E G L P R K L
37034 aaacacgctgcttcttctgtaggcgctgctcgctagatattcaagacgcaggtgaatgggttcgactagtttgcctcct
113 K H V A S S V G A V A L D I I K D A G E W V R L V C P P
37118 gacggtgcttaacaacaagttcgaagcataacagccgcagaaaattcaatgctttggcatcttcgatatcttgcgaagcaaac
141 D G A N K Q V R S I T A A E N S M L W H L R Y L A K Q Y
37202 aatttagaattgacatttgggttatgaagaattatcaagcaagaggttagaattgttcaaacggttgatttcttcagccttat
169 N L E L T F G Y E E I I K Q E V R I V Q T V V F L Q P Y
37286 gtcgagcttaagtagactttcctctgttagttgaagagaatttgaatatgtcactaggcaggaagattctcgaacctgtgt
197 V E S K V D P P L V V E E N L K Y V T R Q E D S R L V C P P
37370 acggttcaagttgacaggttaaaaggaaggaagggcagtcgaagcctttaaagctttgcttctatcaacaatggaaggaatat
225 T A Y K L T G K K E E G S Q E P L T F A S I N N G S E Y
37454 ctcattgatgttctggtttactacacgcccacatgaagcctcgatatattgctaaacttcaaaagcgacgaacattttagaatt
253 L I D V S W F T T R H M K P R Y I A K S K S D E H F R I
37538 aaagaaaatttgatgagtgctgcgctgcttactcttgacatctacagtcgcccactaattggatatagggttcagcggctcctt
281 K E N L M S A A R A Y L D I Y S R P L I G Y E A S A V L C
37622 tataacaaggttctgacttgcatactcaactaattgtcgacgaccattatgatgttatcgagtgccgaaagatatctgct
309 Y N K V P D L H H T Q L I V D D H Y D V I E W R K I S A
37706 cgaaaaattgactacgacgacctttcaaaccttactatcattttccaaagaccctcgaaaagacttgatggacttgaatgag
337 R K I D Y D D L S N S T I I F Q D P R K D L M D L L N E
37790 gacggcgaaagagtcctttcaggggaaactgtaaatgagtcacaggttggatttagatagcagagatgacattttaggggactat
365 D G E G V L S G E T V N E S Q V V I R Y A D D I L G T N
37874 tttaatgcagaatctgggaaatacattgggtgcttctaataactaataagaaaccgagcgaattagttcctgacgactttacatgg
393 F N A E S G K Y I G V L N T N K K P S E L V P D D P T W
37958 attcgactagaaggtcctaaaggtgacgcaggtttaccgggagctcctgggctgagtgagtgacgagtggttacctggaagagc
421 I R L E G P K G D A G L P G A P G R D Q V D G V P G K S
38042 ggatgaggtatagcagatacagctatcattatgctgtatccggtttccggaacgcaagagcctgaaatggatggagcgaacaa
449 G V G I A D T A I T Y A S V S V S G T Q E P E N G W S E Q
38126 gttcctgaaactcataaaaggtcgattcttggactaaaaccttttggagatatactgacggctcacatgaaactggatactcc
477 V P E L I K G R F L W T K T F W R Y T D G S H E T G Y S
38210 gttgcttatataggcgaagacggaattccggaagacggaatcgaggttaaggacggaagtaggtatagccgcaactgaagtc
505 V A Y I G Q D G N S G K D G I A G K D G V G I A A T E V
38294 atgtatgcaagttcgccatctgctactgaagctccagctgggtgagtggtctacgcaagttcctaccgtcccaggttggtcagtat
533 M Y A S S P S A T E A P A G G W S T Q V P T V P G Q Y
38378 ttatggactcgaacaagatggcgctacactgaccaaactgatgaaattggatattcagtttcaagaatggggcagcagggctcctt
561 L W T R T R W R Y T D Q T D E I G Y S V S R M G E Q G P
38462 aaaggtgacgcaggtcgtagcgggtattgcaggaagaaagcgaataggggttgaagtaacttcagtttctttaggaattagtcctc
589 K G D A G R D G I A G K N G I G L K S T S V S Y G I S P
38546 actgattctcgattcctggagatgggttccacaagttctctttaaactcaaggtcaatatcttggactcgaactatttgg
617 T D S A I P G V W A S Q V P S L I K G Q Y L W T R T I W
38630 acctataccgattcaactaccgaaacgggctatcaaaaacctacattccaaaagacgggaatgacggtaaaaatggaaattgct
645 T Y T D S T T E T G Y Q K T Y I P K D G N D G N I A
38714 ggttaaggatggggtaggaattaaagttacgaccattacacgcagggctcaacctcaggaacagttgctgctacttcaaatgg
673 G K D G V G I K S T T I T Y A G S T S G T V A P T S N W
38798 acttctgctattccaaatgttcaaccgggattcttcttggagcgaactgttggaaactatactgatgacactagcgaacaa
701 T S A I P N V Q P G F L W T K T V W N Y T D D T S E T
38882 ggttactcagtttccaaagataggtgaaacaggtccttagaggagttcaaggtcttcaaggtcctcaagggcttcaaggaattcctt
729 G Y S V S K I G E T G P R G V Q G L Q G P Q I P
38966 ggacctgcaggagctgacggagcttcgcaatatactcacctcgcttcttctaagtccaaacgggtgagggatttagtcatact
757 G P A G A D G R S Q Y T H L A F S N S P N G E G F S H T
39050 gacagcggacgagcatagctcggtcagtagtaagatttcaatcccgctccattcaaaagacctgacgctatcacatggagcga
785 D S G R A Y V G Q Y Q D F N P V H S K D P A A Y T W T K
39134 tggaaaggggaatgacggagctcaagggataccgggaagccagggcgacagggtaagactaattatttccatatagtattacgct
813 W K G N D G A Q G I P G K P G A D G K T N Y P H I A Y A
39218 tcaagtgacagcggatcacgtgagttcagtttggagataataatcaacaatatatgggttattactccgattatgaggaagca
841 S A D G S R E F S L E D N N Q Q Y M G Y Y S D Y E Q A
39302 gatagcagggatcgaactaagtagtgagtttgacggccttgccaatgttcaagtgagggtcgaaacgaggttcccttaattctt
869 D S R D R T K Y R W F D R L A N V Q V G G R N E F L N S
39386 ttatttgaatttgggtttaaaccctcgctattctagtacaactcaatggacgggacaagatcaaacgcaaggacagatatctgct
897 C E F F G L K P R Y S Y N L M D G Q D Q T Q I S A
39470 actattgacgaacgtcaacgggttcaaggtgctaacctcttacgacttgactcaacatggaacgggttaaacgcagaaacaaaaa
925 T I D E R Q R F K G A N S L R L D S T W N G K P Q N Q K

```

364

39554 ctgacctttttcttagggagatagcgattaggtactccaaccgagtggtcctaatttagaaggctcgtatcagtttctgggct
953 L T F S L G G D T R L G T P T E W S N L E G R I S F W A
39638 aaggcctctaggaacggagtgagcttagctgcacggccgggttatcgttagtaacgtatttacgcgaacccctaacgcatcaatgg
981 K A S R N G V S L A A R P G Y R S N V F T A T L T D Q W
39722 aagttctacgatttttaattctttgacaaagtttaattcaaatgtaccgctgaagcaattttccatgtattcactcaaaagttgt
1009 K F Y D F K F F D K V N S N C T A E A I F H V F T Q S C
39806 ccagtgtggctcaatcatattaaatcgaacttggtgaatattcttactccttttagtgaagcagaggaaagaccttaaatatcga
1037 S V W L N H I K I E L G N I S T P F S E A E E D L K Y R
39890 attgactcaaaagccgatcaaaagctaaactaaccaacagtgacggcactcacggaagggctcaactacatgacgcagaactg
1065 I D S K A D Q K L T N Q Q L T A L T E K A Q L H D A E L
39974 aaagctaaggctacaatggagcagtttaagtaacttagaaaaggcttatgaaggtagaatgaagcctaaggaagctatcaaa
1093 K A K A T M E Q L S N L E K A Y E G R M K A N E A I K
40058 aaatcggaagccgacctaatcttagcggaagtcgaattgaagctactatccaagaacttgccgggctacgggaactgaagaag
1121 K S E A D L I L A A S R I E A T I Q E L G G L R E L K K
40142 ttcgtcgacagttacatgagctcttctaataaggtcttaattcggttaagaacgacggtagctctaccattaaaggttaaggt
1149 F V D S Y M S S S N E G L I I G K N D G S S T I K V S S
40226 gaccgaattttctatgttctccgcaggggaatgaagttatgtacctacgcaaggggttcattcacatcgataacgggattcttacc
1177 D R I S M F S A G N E V Y L T Q G F I H I D N G I P T
40310 caatccatttcaagtcggccgatttagaagcgaacaatactcgtttaatccagacatgaacgtgattcggtatgtaggataa
40390
1205 Q S I Q V G R F R T E Q Y S F N P D M N V I R Y V G *
dp102P002
32386 atggattttgggtcaattgcagcaaaaatgactttggatattctcaaaactcacaaagtcatttaaatcttgctcaaaagtcgaagcg
1 M D F G S I A A K M T L D I S N F T S Q L N L A Q S Q A
32470 caacggctcgcactagagctcttcgaagtcctttcaaatgggtctgctttaacaggattagggaaggacttacgactgcgggt
29 Q R L A L E S S K S F Q I G S A L T G L G K G L T T A V
32554 accctctcttcttgggtttgagccgctcttataagtagggaatgaattccaagctcaaatgtcttcttcaagctatt
57 T L P L M G F A A A S I K V G N E F Q A Q M S R V Q A I
32638 gcaggagcgacagcggaagagcttggtagaatgaagctcaagcaatcgacctgggtgctaaaactgcttttagtgcgaagag
85 A G A T A E E L G R M K T Q A I D L G A K T A F S K E
32722 gcggctcaaggtatggaaaatctagcttcagccgggtttccaggttaaatgaaatcatggacgctatgccagggggtacttgacctg
113 A A Q G M E N L A S A G F Q V N E I M D A M P G V L D L
32806 gctgcgctatctggagagatgtggccgagctcccgagggccatggctagttcacttcgagcctttggattagagggcaaacccag
141 A A V S G G D V A A S S E A M A S S L R A F G L E A N Q
32890 gcgggtcagctgggtgacgtatttctcgagcagcagctgacgaacgcagaactagcgacatggcagagggcgatgaatac
169 A G H V A D V F A R A A A D T N A E T S D M A E A M K Y
32974 gtcgcaccogttgctcactctatgggttgagcctgaagaacggctgcgtctattgggtatttgccgagcggcggtattaag
197 V A P V A H S M G L S L E E T A A S I G I M A D A G I K
33058 ggctcgcaagccggaaccacgcttagaggcgctctctcgctatttgccaaacctacgaagcgatgggtcaaatcaatgcaggaa
225 G S Q A G T T L R G A L S R I A K P T K A M V K S M Q B
33142 ttaggagtttctgcttcagcagcggaacgggaacatgattccactaagagaacaaatcgctcaactgaaacagctactgcagga
253 L G V S F Y D A N G N M I P L R E Q I A Q L K T A T A G
33226 ctaacacaagaggaacgaatcgctcaccttggtaccttggtatggccaaaactcggtgtcagggtatgcttgactattagacgca
281 L T Q E E R N R H L V T L Y G Q N S L S G M L A L L D A
33310 ggtctgagaaattggataaagatgacaaactgctcgtgaactcggacggagctgctgaaggaaatggcagaactatgcaggac
309 G P E K L D K M T N A L V N S D G A A K E M A E T M Q D
33394 aaccttgcttagtaaaatcgagcaaatgggaggagcttcgagctggtgctattattgttcaacaaatccttgagcctgcactt
337 N L A S K I E Q M G G A F E S V A I I V Q Q I L E A P L
33478 gctaaaatcggtgggagcaatcacaaagttctcgaagcattcgtaaatatgtccactatcggtcaaaagatgggtgtcatattc
365 A K I V G A I T K V L E A F V N M S P I G Q K M V V I F
33562 gcaggaattgttcagcccttggaaccactgcttcaattgcaggaatgggtgatgacaactattgtcaagtttaagaattgctatt
393 A G M V A A L G P L L L I A G M V M T T I V K L R I A I
33646 cagtttttaggtccagcatttatgggaacgattgggaacattgcaggagttatagcaatattctatgctctggtcgcgctgttc
421 Q F L G P A P M G T M G T I A G V I A I F Y A P V A V F
33730 atgatagcctacacaaatcgagagatttagaaactttatcaacagctcttgccgctgctattaaagctgggtttggaggagcg
449 M I A Y T K S E R F R N F I N S L A P A I K A G F G A
33814 ttggaatggctacttccagcagtaaaagagtttagagaatgggttacagaaggcaggcgagaaggcgaaagagttcggtcagttc
477 L E W L L P R L K E L G E W L Q K A G E K A K E F G Q S
33898 gtagggtctaaagtgctaaaactgctcgaacagtttggaataagtatcggtcaggcaggaggtcgattggtcagttcattgga
505 V G S K V S K L L E Q F G I S I G Q A G G S I G Q P I G
33982 aatgttctcgaaaggttaggagggcgatcttggaagtaggagagtcatttcaattgctgtttcacttgtaacaaaattcggt
533 N V L E R L G G A F G K V G G V I S I A V S L V T K F G
34066 ctgcgatttctagggtattacaggaaccactcgggattgctattagttctgttagtttcttttgacagcttgggttagaacaggt
561 L A P L G I T G P L G I A I S L L V S F L T A N A R T G
34150 gagttcaacgcagcgaattactcaagttatcgaaaacttgacaaacacaaatcagtcgacggctgatttcttcttcaatc
589 E F N A D G I T Q V F E N L T N T I Q S T A D F I S Q Y
34234 cttccagctcttctcgaaaagggaactcaaattttagttaagattattgaagggaattgcacatgctgttctcctaagtagtgaa
617 L P V F V E K G T Q I L V K I I E G I A S A V P Q V V E
34318 gtgatttcacaagtcattgaaaatattgtgatgacaatttcgacagttatgctcatttagtcgaagcagggaattagaatctt
645 V I S Q V I E N I V M T I S T V M P Q L V E A G I K I L
34402 gaagcgcttataaatggtcttcttcaatctcttcttcaatcattcaagcagctggtcaaatctcactgcttcttcaatggt
673 E A L I N G L V Q S L P T I I Q A A V Q I I T A L F N G
34486 cttgttcaggcacttctacgcttattcaagcagctcttcaaatttgtcagctctcataaacggactagttcaagcgcttccg
701 L V Q A L P T L I Q A G L Q I L S A L I N G L V T A L N P
34570 gcaattattcaagcagctggtcaaatctatcgtcgcttgctcaagcactaattgaaaacttgcttctgataatcgaaagcagcg
729 A I I Q A A V Q I I M S L V Q A L I E N L P M I I E A A

34654 atgcagattataatgggtctagtcacgcactgattgaaaatataggacctatcttagaagcagggtattcaattcttaattggct
757 M Q I I M G L V N A L I E N I G P I L E A G I Q I L M A
34738 ttaatcgagggtacttattcaagtgtctctgaactaattacagcagcgattcaaatcattacttctactattagaagcaattcttg
785 L I E G L I Q V L P E L I T A A I Q I I T S L L E A I L,
34822 tcgaaccttctcaacttctagaagccggagttaaattgcttttatcacttcttcaagggttgctaaatagcttctctcaacta
813 S N L P Q L L E A G V K L L L S L L Q G L L N M L P Q L
34906 attgcaggggtcttgcataatcatgatggcacttcttaagcagttatcgacttctgctccctaaacttcttcaagcaggtgttcaa
841 I A G A L Q I M M A L L K A V I D F V P K L L Q A G V Q
34990 cttcttaaggcattgattcaaggattgtcttcaacttctcggtcacttttatcgacagctggaaacattgcttctcatcattagt
869 L L K A L I Q G I A S L L G S L L S T A G N M L S L V
35074 agcaagattgctagcttctgaggacagatgggttcaggaggtgcgaacctgattcgaaacttcttagtggtattgggtcaatg
897 S K I A S F V G Q M V S G G A N L I R N F I S G I G S M
35158 attgggtcagctgtctctaaaattggcagcatgggaacttcaattgttcttaagggttactggattcgctggcaaatggtaagc
925 I G S A V S K I G S M G T S I V S K V T G F A G Q M V S
35242 gcaggggtcaacttctgctgaggatttatcaatggatcagttccatggtaagttctgcggttaagtcggcggtcaataggct
953 A G V N L V R G F I N G I S S M V S S A V S A A N M A
35326 agcagtgcatataatgccgttaagggtattcttaggtattcacttctctcagctgtcatggagcagatgggtatctatacgggt
981 S S A L N A V K G F L G I H S P S R V M E Q M G I Y T G
35410 caagggttcgttaaatgggtattggttaacatgattcgaaactacacgtgacaggtcaaaagaaatgggtgaaactgttactgaagct
1009 Q G F V N G I G N M I R T T R D K A K E M A E T V T E A
35494 ctcagcgacgtgaagatgatatcaagaaaatggagttatagaaaagggttaaatcagtttacgaaaagatggctgacacact
1037 L S D V K M D I Q E N G V I E K V K S V Y E K A M A D Q L
35578 cctgaaacttctcagctcctgatttgcgaagatgttcgtaaaagcagcgggttcgctcagtgaggacttcttcaatcacaggaagt
1065 P E T L P A P D F E D V R K A A G S P R V D L F N T G S
35662 gaacaccttaaccaactcagtcacaaactcaaaaacacacagcgagcaaacctgttgaacattggaaacattcgtatttga
1093 D N P N Q P Q S Q S K N N Q G E Q T V V N I G T I V V R
35746 aacaatgacgagcttgacaaactgtcgagaggtatgtatagaagtaaaagaaacttctcaggggttggtaacattgttaaca
1121 N N D D V D K L S R G L Y N R S K E T L S G F G N I Y T
35830 ccgtaa 35835
1149 P *
dp1ORF003
53538 atggcacaaaaaggactctttgggtgcaagcctcgttctagcaagaagaacgatgctcagttacttggctcaacggaaaaaacagg
1 M A Q K G L F G A K P R S S K K N D A Q L L A Q R K N R
53622 aagcctgcagttgaggttacttactattcaggaacgctctaaaggacgcaggttgcagagctcgtactcttcaactaggatt
29 K P A V E V T Y I S G N A L K D A V A R A R T L S T R I
53706 cttggacacgttcttgatagacttgagtttaactcagtaggaagcaaaactcgagcagtagtagacaaaatgattgaagacgga
57 L G H V L D R L E L I T E E A K L E Q Y V D K M I E D G
53790 atagggttctattgacgttagaaactgatgactcgatattcagcatgagctggcaggagctcgttctgactcactcagttcaa
85 I G S I D V E T D G L D T I H D E L A G V C L Y S P S Q
53874 aaaggaatctgctcctgtcaatcatgttgaatatagcaagatgcgaatgaagaatcaaatcttctcctgagttcatgaag
113 K G I Y A P V N H V S N M T K M R I K N Q I S P E F M K
53958 aaaaatgctcaacgattgtatagattcaggaattcctgtcatctatcataattcgaaatttgacatgaaactgatttattggcga
141 K M L Q R I V D S G I P V I Y H N S K F D M K S I Y W R
54042 ctcggcgtcaaaaatgaatgagccagcgtgggtatcatatttagccgcaatgcttttaaatgaaaacagagttctcacagctgaaa
169 L G V K M N E P A W D T Y L A A M L L N E N E S H S L K
54126 agtcttacttcaaatatgttaggaacgaagaaaacgcagaggttgcaaaatttaatgacttatttaaggaattccttttagt
197 S L H S K Y V R N E E N A E V A K F N D L F K G I P F S
54210 ttaacttctcctgacgttgcctatgtatgcggcctatgaccccttgcgaacttctgaactctgaattcagaacaaactac
225 L I P P D V A Y M Y A Y D P L Q T F E L Y E F Q E Q Y
54294 ttgactccaggaaactgaacaaatgtgaagaatataaactggaaaaagttctcatggttcttcaaatattgagatgccttcaat
253 L T P G T E Q C E E Y N L E K V S W V L H N I E M P L I
54378 aaagtctctctcgacatggaagtctacggtgtcgacttagaccaagataagctggcagaatttagagaacagtttactgccaat
281 K V L F D M E V Y G V D L D Q D K L A E I R E Q F T A N
54462 atgaacgaggtgagcaagagtttcaacagcttgcagcgaatggcagcctgaaattgaagaacttcgacaaactaatttccag
309 M N E A E Q E P Q Q L V S E W Q P E I E E L R Q T N F Q
54546 agctatcaaaaactcgaaatggatgcaagaggtcgagtgacggttaagcatttccagctctactcaattagcaattctgtttat
337 S Y Q K L E M D A R G R V T V S I S S P T Q L A I L F Y
54630 gatatcatgggtatgaaaagtcctgaaagggataaaccttagaggaacagggcgaaggtattgtcgagcattttgataacgatatt
365 D I M G L K S P E R D K P R G T G E S I V E H F D N D I
54714 tcaaaagcacttttgaatatagaaaatatgcaaaatttagtttcgacctatacaacacttgaccaaacaccttgcaagcctgac
393 S K A L L K Y R K Y A K L V S T Y T T L D Q H L A K P D
54798 aatcgaattcacactacattcaaacagtcaggagctaagacagggcgtatgtcaagtgaagaatcctaactacagaatattcct
421 N R I H T T F K Q Y G A K T G R M S S E N P N L Q N I P
54882 tctcgcggtgaggtgcagtagttcgacaaaactcttcagccagtgaaagggcattacattattggtagtgcactctcaacaa
449 S R G E G A V V R Q I F A A S E G H Y I I G S D Y S Q Q
54966 gaacctcgttcttggcggaatttaagtgccgacgaagatgcgacatgcttcaacaaaaacctggacctatattcagttatc
477 E P R S L A E L S G D E S M R H A Y E Q N L D L Y S V I
55050 gggtcgaaactttatgggttccctatgaagagtggttagagttctatccgcaggaacgactaacaaggaaggaacttcga
505 G S K L Y G V P Y E E C L E F Y P D G T T N K - E G K - L R
55134 agaaattctgtcaagtcggttcttttaggtcttctgtcggccgggggtaactcaatcgctgagcagatgaatgtatctgtc
533 R N S V K S V L L G L M Y G R G A N S I A E Q M N V S V
55218 aaagaagcgaataagggtattgaagatttctcaccaggttccctaaagtggcagactatcatattcgttcaacagcagggcg
561 K E A N K V I E D F F T E P P K V A D Y I I F V Q Q Q A
55302 caggacttgggtatgttcaaacagctacgggtcgaaagaggttctctgatagagttctcctgaatacagagttcgagtat
589 Q D L G Y V Q T A T G R R R R L P D M S L P E Y E F E Y
55386 atcgacgctagcaagaacgaagatttcgaccccttaactttgacgcagaccaacagatggacgatactgttctgacatatt

366

617 I D A S K N E D F D P F N F D A D Q Q M D D T V P E H I
55470 atcgaaaaatattgggcccagctagatagagcctggggtattaaagaagaagcaagaattaaagaccaggcaaaagccgaagga
645 I E K Y W A Q L D R A W G F K K K Q E I K D Q A K A E G
55554 attcttattaaggataacggaggcaagatagctgatgctcagcgccaatgtttgaactcagttattcaaggaaacggcagccgac
673 I L I K D N G G K I A D A Q R Q C L N S V I Q G T A A D
55638 atgactaagtacgcaatgattaagggtacacaatgacgctgaattgaaagaattaggattccatttaattgattccagttccagat
701 M T K Y A M I K V H N D A E L K E L G F H L M I P V H D
55722 gagttactagggtgaggttccatcaagaacgcaaaacggggagcagaagggtgacagaagttatgattgaagcagcgaaggac
729 E L L G E V P I K N A K R G A E R L T E V M I E A A K D
55806 attattagcttccaatgaaatgtgacccagctatagtagaagatgggtatgggtgaagaaattgaaatctaa 55877
757 I I S L P M K C D P S I V E R W Y G E E I E I *

dp10R7004
40401 atgcacaaatattcaactcatagcgccctcttccacttgaacctttacgtcgaacaagtttagtcaggacgtaacgaacaactcc
1 M T K F I N S Y G P L H L N L Y V E Q V S Q D V T N N S
40485 tcgagagtttagttggcgagctactgtcgacccgagatggagcttatcgaaactgggacttatggaaaatttagtaacctttccgta
29 S R V S W R A T V D R D G A Y R T W T Y G N I S N L S V
40569 tgggttaaatgggtcaaggttccatagcagtcacccagactacgacagctccggcgaaggttaacgctcgcaagtggaaggtg
57 W L N G S S S V H S S H P D Y D T S G E E V T L A S G E V
40653 actgttccctcacaatagtgacgggacaaagacaatgtccggttgggcttcggtttgacccctaataacggcggttcacggaaatctc
85 T V P H N S D G T K T M S V W A S F D P N N G V H G N I
40737 actatctctactaattacacttttagacagctattccaaggtctacacagatttctagttttgagggaattcgaaatctaggatct
113 T I S T N Y T L D S I P R S T Q I S S F E G N R R N L G S
40821 ttacatacgggttatctttaacggaaaagtgaactctttacgcacatcaagtttggtaacgagttttcggtagcgactggatagat
141 L H T V I F N R K V N S F T H Q V W Y R V F G S D W I D
40905 ttaggtaagaaccatactactagcgtatcttccacggcgtcactggacttagcaaggtacttaccataaactcaagttccggaaca
169 L G K N H T T S V S F T P S L D L A R Y L P K S S S G
40989 atggacatctgtattcgaaactataacggaaactacgcaaatggtagtgacgtctattcaaacggatggaggttcaacatcccc
197 M D I C I R T Y N G T T Q I G S D V Y S N G W R F N I P
41073 gattcagtcagctcctacttttccgggcatttcttttagtagacacagacttcagcggttcgacagattttaacagggaacaacttc
225 D S V R P T F S G I S L V D T T S A V R Q I L T G N N F
41157 ctccaaatcatgtcgaaacttcaagtcacttcaacaatgcttccggcgcttaccgattccactatccaaagcatttcacgctgag
253 L Q I M S N I Q V N F N N A S G A Y G S T I Q A F H A E
41241 ctctaggttaaaacccaagctatcaacgaaaacggcgcaaatgggtatgatgaactttaatggctccgctaccgtaagagca
281 L V G K N Q A I N E N G G K L G M M N F N G S A T V R A
41325 tgggttacagacacgagggaaaacaatcgaaactccaaagacgtatctatcaatgttatagaactacttgacgcttatcaat
309 W V T D T R G K Q S N V Q D V S I N V I E Y Y G P S I N
41409 ttctcgttcaacgtactcgtcaaaatcctgcaaatccaagctcttcgaaatgctaaaggtcgacacataacggtaggaggt
337 F S V Q R T R Q N P A I I Q A L R N A K V A P I T V G G
41493 caacagaaaaacatcatgcaaatctacttctccggtggcggttgaaacactactaatttcacagaagatagaggttcggogtca
365 Q Q K N I M Q I T F S V A P L N T T N F T E D R G S A S
41577 gggacggttcaactatttccctaagtactaactcgtccgcaacttagctggtaactacgggcccgaacaggttcaactaggt
393 G T P T T I S L M T N S S A N L A G N Y G P D K S Y I V
41661 aaggctaaaaatccaagacaggttcaacttcgactgaatttagtgcacggttagctaccgaatcagtagttcttaactatgacaag
421 K A K I Q D R F T S T E F S A T V A T E S V V L N K
41745 gacggtcgacttgaggttggttaagggttcagaacaagggaagggcaggtcaattgatgcagcaggtgatataatgctggaggt
449 D G R L G V G K V V E A Q G K A G S I D A A G D I Y A A G G
41829 cgcaaggttcaacagtttcagctcactgataataatggagcattgaacaggggtcaatataacagatgtttggaataagcgtgaa
477 R Q V Q Q P Q L T D N N G A L N R G Q Y N D V W N K R E
41913 acagagtttcatggcgaagtaacaaatcagaggacaacctacgggaactcgaggtgaatggggactatttcaaaatttctgg
505 T E F T W R S N K Y E D N P T G T R G E W G L F Q N F W
41997 tttagatagctggaaaatgggttcaatcttctacataatgtcaggaagaatgttcacagggacagcgaacgtggaacagctgg
533 L D S W K M V Q S F I T M S G R M F I R T A N D G N S W
42081 agacctaaacagtggaagaggttctatttaagcaagacttcgaacagaataattggcagaacttgttcttcaagtggggtgg
561 R P N K W K E V L P K Q D F E Q N N W Q K L V L Q S G W
42165 aaccatcactcaacctatggcgacgcttcttctgaaaactcttgacggcatagtatatttgagagggaatgtgcataaagga
589 N H H S T Y G D A F Y S K T L D G I V Y L R G N V H K G
42249 cttatcgacaaagaggttactattgcagtaacttctgaaggatttagaccgaaagtttcaatgtatcttcaggctctcaataac
617 L I D K E A T I A V L P E G F R P K V S M Y L Q A L N N
42333 tcatatggaatgccattctatgtatatacactgacggaagacttgggtgaaatcgaaatgtagataattcttgggttaattta
645 S Y G N A I L C I Y T D G R L V V K S N V D N S W L N L
42417 gacaatgcttctatttctgatttaa 42440
673 D N V S F R I *

dp10R7005
23674 atggctaaaaaatcaaaagctattctcacacacagcgaactgattagtcagtcggtttgacagcccttggcagaagaatcaaaag
1 M A K K S K A I S H T D E L I S Q S P D S P L A K N Q K
23758 ttcaagaagagcttcagggaagttgaaaagtattatcaacttcgacggatgtgacggaacttgaatactgactatggg
29 F K K E L Q E V E K Y Y P D G P D V T D L N T D Y G
23842 caaacatggaagattgacgaagactcagtcgactataaacctactcgagaatctcgaaactatattcgacaacttatcaaaag
57 Q T W K I D E D S V D Y K P T R E I R N Y I R - Q L I - K K
23926 caatcagctttatgatgggttaagagcagagcttctttagtccagttcaagacaatcaagatgaacaggtcgagaacaag
85 Q S R P M M G K E P E L I F S P V Q D N Q D E Q A E N K
24010 cgtattcttctcagctctattttaaggaattgttaattctggagcaaaagtacaaatgcatttagtcgacggccacagtaggtaag
113 R I L F D S I L R N C K F W S K S T N A L V D A T V G K
24094 cgggttattgatgacagtagtagcaaatgccgtcaacaaattgacgtccagttttattcaatgcctcagttcacctatacagtt
141 R V L M T V V A N A A Q Q I D V Q F Y S M P Q F T Y T V
24178 gaccctagaacccctccagcttcttctgttgcattgtttatcaggacgagcgtacaaaaggaatgagcactgaaaaacaa

367

169 D P R N P S S L L S V D I V Y Q D E R T K G M S T E K Q
24262 ctttggcatcattatagatatgaaatgaaagctggaacaaagtcacaggaattgcaacagcttagaagacattgaagaacaa
197 L W H H Y R Y E M K A G T S Q S G I A T A L E D I E E Q
24346 tggctggctcacttatgccttaacggatggagagtcgaaccaaactctatatgacagaaagtgccaaactcttatcaaggagaca
225 C W L T Y A L T D G E S N Q I Y M T E S G Q T T I K E T
24430 gaggtcaaaacttgtagaaattgaagacaacctaggaacaaagattgaagttcctttaaaagttcaagaatccgcccacccggc
253 E A K L V E I E D N L G N K I E V P L K V Q E S A P T G
24514 ttgaagcaaaactccttgctgagttattcttaatgaaccattgactaatgacatatagggacaagcagtgcaaaagaccttatc
281 L K Q I P C R V I L N E P L T N D I Y G T S D V K D L I
24598 acagtagcagataaacttgaaacaaactattagtgacttacgagattcacttcgattttaaagtgtcgagcagctgttatcatt
309 T V A D N L N K T I S D L R D S L R P K M F E Q P V I I
24682 gatggctcttcttaagtcattcaaggaaatgaagattgcccacacgcttggctgaccttaagagtgacaccttcttctcaatc
337 D G S S K S I Q G M K I A P N A L V D L K S D P T S S I
24766 ggccgacttgaggcaagcgaagcgaagtcacttccttcaggaactcacttccttcagcggctgaattattatttagag
365 G G T G G K Q A Q V T S I S G N P N P L P A A E Y Y L E
24850 ggccgtaagaaagccatgtagaactaatggacagccaatgcctgaaaggtacaggaggcgccatcaggaattgcaatgcag
393 G A K K A M Y E L M D Q P M P E K V Q E A P S G I A M Q
24934 ttcttattctacgacctaatttctcgatgtgacggaaaatggattgagtgaggatgctattcaatggctcattcaatgctg
421 F L F Y D L I S R C D G K W I E W D D A I Q W L I Q M L
25018 gaagaaatcttagcaacagtgaaatggtgacttgggaaacttctcaagataattcaatcaagttatcaaaactacagcaatg
449 E E I L A T V N V D L G N I P Q D I Q S S Y Q T L T T M
25102 actatcgaaacaccactatccaattcctagcgatgaacttctcgtcaagcaacttgcgctcactgaagttcaaaactaatgtacgc
477 T I E H H Y P I P S D E L S A K Q L A L T E V Q T N V R
25186 agcccaactcttacttgaaagattcagtaagaaggaagggcggaaggaatgggaacgcatttggaaagaaacttgcctcag
505 S H Q S Y I E E F S K K E K A D K E W E R I L E L A Q
25270 cttgacgaatctcagctggagcattgcctgtattagcaaacgaattaaacgaacaagaggagcctcaagatgaacagcagtgaa
533 L D E I S A G A L P V L A N E L N E Q E E P Q D E T S E
25354 gaagacgaagttgatgacaaagaaagaacaaactgaacaacacccaaggaaggaagtcgaccacagcttcaagggttaa
25434
561 E D E V D D K E K E Q T E Q P T E E G V D P D V Q G *
dp1ORF006
45296 atgattgaatcggttatagcagcttcgaaagctaggcgaggtcgaaacctatttattgaacatggggcaagcactgatgaagat
1 M I E I V I A R S K A R R G R T L F I E T W A S T D E D
45380 gcagttaaaatggcagaaaagattccagcttcccaatgtagtcgagagcttcttaataacttcgaactaccttaccattatagat
29 A V K M A E K I S S L P N V V E T S S N N F E L P Y K Y
45464 ttcaataatgttatagacgctctagatgaatggagcttcacatcttcggcgaacttgataaagatgttcaagatcaattgac
57 F N N V I D A L D E W E L H I F G E L D K D V Q D Y I D
45548 tctcgaaacccaatagcttcttcaagcaatgagcagtttctgcttcaagactactccattcgcgaccaggttgatgttctgaa
85 S R N R I A S S S N E Q F S F K T T P F A H Q V E C P E
45632 tacgcacaagagcatccatgttctcttttaggcgagtagcaaggttttagggaaaactaaacaggcaattgatattgcagtttagc
113 Y A Q E H P C F L L G D E Q G L G K T K Q A I D I A V S
45716 aggaaggcaagtttcaaacattgtttaatcgatgtgtgcataatcagggctcaaatggaaattgggcaaaagagtaggtattcat
141 R K A S F K H C L I V C C I S G L K W N W A K E V G I H
45800 tcaaatgagtcagctcatttttaggaagtcagagtcactaaagatgggaaattagtgattgacggagtttctaaacgggcagaa
169 S N E S A H I L G S R V T K D G K L V I D G V S K R A E
45884 gacttgcttgggtggccacgacgaattcttcttatcactaacattgaactcttcgcgagtgctgtgttcaataacttaaat
197 D L G G H D E F F L I T N I E T L R D A V F I K Y I D
45968 gaactgacaaaaagcggagaaatggaaatggttattattgacgagattcacaagtgaagaaccttcaagtaagcaaggggct
225 E L T K S G E I G M V I I D E I H K C K N P S S K Q G A
46052 tcaattcaaaagctccaaagttattacaagatgggacttacaggaactcctcctaatgaataacccaatcgatgtattcaargtt
253 S I Q K L Q S Y Y K M G L T G T P L M N N P I D V F N V
46136 atgaagtggttagggcggaacatcacactgacttgaagagcgatgactgtatcgtcgaccagttcaatcaaatcact
281 M K W L G A E H H T L T Q F K E R Y C I V D Q P N Q I T
46220 ggatattcgaaatctagctgaacttcgcgagcttgtcaacgactacatgcttagaagaacgaaggaaggttttagacctgcct
309 G Y R N L A E L R E L V N D Y M L R R T K E E V L D L P
46304 gaaaagattcgagtcacagagtgatgtcgacatgaactcgaacagtcacaaactctataaggaagttttgactaaacttgttaa
337 E K I R V T E Y V D M N S K Q S K I Y K E V L T K L V Q
46388 gaaatagataaagtcaggtcattgcctaacctcctagccgaaacgattcgacttcgacaagcagctggaaatccttcgatttta
365 E I D K V K L M P N P L A E T I R L R Q A T G N P S I L
46472 actactcaagatgtcaagcttgcgaagtcgaagatgtatcgaaattgtcgaggaatgtatccagcaaggaaagtcctgcgtg
393 T T Q D V K S C K F E R C I E I V E E C I Q Q G K S C V
46556 atatttagcaattgggaaaaggttattgaaacctcttgctcaagatacttgcgaagacagtcacaaatgcacactggtaacagagaa
421 I F S N W E K V I E P L A K I L S K T V K C N L V G E
4640 accgcagataagttcaacgaattgaagaatttatgaatcacagaaggcttctgtatttttaggaactataggtgcgctagga
449 T A D K F N E I E E F M N H R K A S V I L G T I G A L G
46724 acaggatttactttgacgaaagcgatagcttatttcttagatagtcggtggacgcgcagagaaagcaccagccgaagat
477 T G F T L T K A D T V I F L D S P W T R A E K D Q A E D
46808 aggtgtcatagaattggcgcaaaagttctgtcactatctacagcgttgcgcgaaggtactgttgacgaacgtatagaagac
505 R C H R I G A K A S S V T I Y T L V A K G T V D E R I E D
46892 cttattgaacggaaaggagaattagcagatttatcgtagatggttaagcctatgaatctaaattggtaacctttcgatattc
533 L I E R K G E L A D Y I V D G K P M K S K I G N L F D I
46976 ctgcttaaatag 46987
561 L L K *
dp1ORF007
22230 atgacaataagcctgagaaataaactacctaagttcaacttcgctccttttagtaagaaacaactccagctcctaactatggctgg

1 M T I S L R N K L P K F N F V P F S K K Q L Q L L T W W
22314 acaaaagggctcaccttttcgaactttcgatattcgatagcagagcggttccattcgttcaggaaaaacagratcgatggctctt
29 T K G S P F R T F D I V I A D G S I R S G K T V S M A L
22398 tcattttccctttgggccaatgacggaattcaacggacaaaactttgccattctgtgtaagacattcactcagctcgacgaaat
57 S F S L W A M T E F N G Q N F A I C G K T I H S A R R N
22482 gttattcagctctctaaagcaaatgctcacaagtcgcgggtatgaaattcgagatgttcgaaatgaaatctacttattattaga
85 V I Q P L K Q M L T S R G Y E I R D V R N E N L L I I R
22566 cactttagaatggcgaagaattgtcaactacttctatatatttggaggaaaagatgagtcgagtcgaaccttatacagggg
113 H P R N G E E I V N Y F Y I F G G K D E S S Q D L I Q G
22650 gtaacattagcaggtatcttctgtgatgagggtggcactgatgcctgaatcgtttgcacccaagcgacggcgctgttccgta
141 V T L A G I F C D E V A L M P E S F V N Q A T G R C S V
22734 acaggttcgaaaatgtgggtcttctgttaaccggcccaatcctactacttcaagaagaactggattgacaaaacaggtcgaa
169 T G S K M W F S C N P A N P N H Y F K K N W I D K Q V E
22818 aagcgtatcttatcttactttacaatggacgacaaaccttagcttgacggatagcattaaaaaggcgctatgagaaaatgtat
197 K R I L Y L H F T M D D N P S L T D S I K R R Y E K M Y
22902 gctggagcttcaggaaaagatttattctcgcccttgggttaacagcagatggcttagtttattcaatgttcaatgaagagcag
225 G A V F R K R F I L G L W V T A D G L V Y S M F N E E Q
22986 catgtcaaaaagctcaatatagaattcgacgggttattctgtacgagcgacttgggtatctataatgcaacaaccttcggcctt
253 H V K K L N I E F D R L F V A G D F G I Y N A T T F G L
23070 tatggattctcgaacgctcataagcgctaccatctaattgagtcatactaccactcagggcgcgaggcggaagagcaactaact
281 Y G F S K R H K R Y H L I E S Y Y H S G R E A E E Q L T
23154 gagcggtatgttaattcgaaatattcaattttagttcagttctacaaaagactactaaagagtcagcgaatgatttagtcgatag
309 E A D V N S N I Q P S S V L Q K T T K E Y A N D L V D M
23238 atacgagaaaagcaaatcgaaatataattctcgaccgctctgcttctgctatgattgttgaacttcaaaagcatccttatata
337 I R G K Q I E Y I I L D P S A S A M I H V E L Q K H P Y I
23322 gctagaagaatatacctatcattcctgctcgaatagcgtgacgcttggcatttcatttcacgctgaactcctggctgagaat
365 A R K N I P I I P A R N D V T L G I S F H A E L L A E N
23406 agatttcaactcgacctagcaacacgcagcagcattgaataactatgcttacagctgggacagtaaacgagcgaacggga
393 R F T L D P S N T H D I D E Y Y A Y S W D S K A S Q T G
23490 gaagatagagtcattaaagagcatgaccactgcattggataggaacagatatgcctgtctcactgacgctctaataacgatgac
421 E D R V I K E H D H C M D R N R Y A C L T D A L I N D D
23574 ttcgggttcgaaatcaaatattatccggaagggcgctagaactaa 23621
449 F G F E I Q I L S G K G A R N *
dp1ORF008
49624 gtgatacagcttcaagctcttaataaagttctcgaagaaaagagcttatccattttagaataaattggaattgaccagaatac
1 V I Q L Q V L N K V L E E K S L S I L E N N G I D Q B Y
49708 ttcacggattatttagacgagtatcaattttatcaagaaacttttcgagataggaagagttccggacgacgaaactattctc
29 F T D Y L D E Y Q F I Q E H F S R Y G R V P D D E T I L
49792 gaccattttcctggattcgaatttttcgaatggcgaaactgatgaataccttatcgacaagctaaaagaggagcatctat
57 D H F P G F E P F E I G E T D E Y L I D K L K E H L Y
49876 aattcactgttccaatttttaacggaagcggtcaggacattcaagtagatagtaacattgagcattcggaatataattccaaaa
85 N S L V P I L T E A A E D I Q V D S N I A I A N I P K
49960 ctagaagaacttttcaatcgcttcaattcgttaggggactagacattgctcgaatgctaaacttcgactagactggcggaat
113 L E E L F N R S K F V G G L D I A R N A K L R L D W A N
50044 actattagaaccatgacggtgaagacttgaattatcgacaggttgaactattggacgagctgtggaggcttactctct
141 T I R N H D G E R L G I S T G F E L L D D V L G G L L P
50128 ggtgaggtattgattgtcataatggctcgactggcaggaagtaagtcgtggactattgataaaatgcttgaactgcttgaag
169 G E D L I V I M A R P G G K S W T I D K M L A T A W K
50212 aacgggcatgatgctctctatatagcggggaatgagtgaaatgcaagtgtggtgctcgtatagatactattcttccgaatgtt
197 N G H D V L L Y S G E M S E M Q V G A R I D T I L S N V
50296 agcatcaattcaattaccaagggatttggaaacgacattcagttcgaataatagaggacattatcaagcaatgactgaggt
225 S I N S I T K G I W N D H Q F E K Y E D H I Q A M T E A
50380 gaaaattcccttgggtgagtcacgcccccttatgattggggaagaaaccttaccctcgcaatttttagatagcatgatctaaa
253 E N S L V V V T P F M I G G K N L T P A I L D S M I S K
50464 tatagaccatctgtggtggggtgaccagctttcactcatgagcagcttattccaagcaggagcagaagcgaatccagtac
281 Y R P S V V G I D Q L S L M S E S Y P S R E Q K R I Q Y
50548 gccacatcaccatggacctatataagatttctgctaaataggaattcctattgtgcttaattgtccaagcagggtcttggct
309 A N I T M D L Y K I S A K Y G I P I V L N V Q A G R S A
50632 aaaactgaagcgctgaaagtatggaactagaacatatacgagaagtgatggagtaggtcaaatgctagcagagttatcgct
337 K T E G A E S M E L E H I A E S D G V G Q N A S R V I A
50716 atgaagcgtgacgaaaaatccggcatacttgaactatctgctcgttataaaacccgataggcgaagaccgaaaaatcatcgaatat
365 M K R D E K S G I L E L S V V K N R Y G E D R K I I E Y
50800 atgtgggacgttgaactggaacctatactctttaggttcaaaagaggagcgaaggaactgaaaagggcgaagctct
393 M W D V E T G T Y T L I G F K E E G E E G T E K G E S S
50884 ccattgaaagcaaaagcctctaggtcgactgctcgtctcgaagttaggttacaaggggaaggttgaagcattttga 50961
421 P L K A K A S R S T A R L R S K V T R E G V E A F *
dp1ORF009
13160 atgacagacttttaaaaacgcttcaagaagcagtaaacgaaacaatcaatcgtagcggtatcgagaaccttatggattggctc
1 M T D F K K R F K K A V T E T I N R D G I E N L M D W L
13244 gaaaatgataccaatttcttctcaagttccagcaagcactcgataccatggaagctatgaaggtggactgttcgagcactatta
29 E N D T N F F S S P A S T R Y H G S Y E G G L V E H S L
13328 aacgtgttcaatcaactacttttcgaaatggataccatggtaggcaagggctgggaagacatttaccctaaggaacagttgca
57 N V F N Q L L F E M D T M V G K G W E D I Y P M E T V A
13412 atcgtagcactatttcacgacctttgcaaagttggcagtatcgtgaaactgaaatggcgcaagacagcagcggtgaatgg
85 I V A L P H D C K V G Q Y R E T E K W R K N S D G E W
13496 gaaagctatttagcatatgaatacagacctgagcaactacaatgggacatgggtgcaaaatctaatttcttcttcaacgttcc

113 E S Y L A Y E Y D P E Q L T M G H G A K S N F L L Q R F
13580 attcaactcagccagttgaagctcaagcaattttctggcatatgggagcctatgatattagtccttatgcaaaatttgaatgga
141 I Q L T P V E A Q A I F W H M G A Y D I S P Y A N L N G
13664 tgtggagcagccttcgaaactaatccatttgcattcttaattccatcgccagatattggccgcaacttatgtatgtagtgaataatgaa
169 C G A A F E T N P L A F L I H R A D M A A T Y V V E N E
13748 aacttcgaataactctcaaggtccagttgaacaagaggctgaggttgaagaagtagttgaagaaaaacctaaagagttcaactcgt
197 N F E Y S Q G P V E Q E A E V E E V V E E K P K S T R
13832 aagaaacctcgccctaaggaagaaaaagttgaagaggctgaagaaaaacaaaagctggaatcactcgagctcgcaaacctcgcg
225 K K P A P K E E K V E E A E E K P K A G I T R R R K P A
13916 ccaaaagaggaagaggttagaagagcctaagaagagccttaagaagcattcttcaaaattcgaaatgcctaaaaagactgaaaaag
253 P K E E S V E E P K E E P K K A S S K I R M P K K T E K
14000 gtcgaagaggttagaagcgcagacgagccgaaagttgaagaagcagaggacgacaatgtggtggtacctgctggatattgttcga
281 V E E V E S A D E P K V E E A E D D N V V V P A G Y V R
14084 gatgtctactacttctacagtgaagtcgctgacgtttactacaagaagatgtcgacgagcctgacgatgacagcgacattctt
309 D V Y Y F Y S E V A D V Y Y K K D V D E P D D D S D I L
14168 gtacaggaagagtagatgagcgcagcagcgaagttgaagaagcagaggacgacaatgtggtggtacctgctggatattgttcga
337 V D E E E Y M D A M C P V L E E D F F Y E L D G K V H K
14252 ttacgaaaggtgaacgcttgcgcgaagaatcagcgaagaacacttgggaacattcactgaagcagaatcactcaagcgaaca
365 L A K G E R L P E E Y D E E T W E P I T E A E Y I K R T
14336 gaaaaacctaaagcagttgcaaacctactcgaaaaactccagcgcttctcgtcgccctcgcccttaa 14404
393 E K P K A V A K P T R K T P A P S R R P R P *

dp1ORF010

8699 atgaaattggaacagttgatgaaggactggaataaggattcgaaagctctttagtagcagttcaaggacttgaacgtgaagcgctt
1 M K L E Q L M K D W N K D S K A L V A V Q G L E R E A L
8783 ccaagaatcccttttctgcgccttctatgaattatcaaaacctacggcggtccctcgaaaaagggtagttgaattcttcggt
29 P R I P P S A P S M N Y Q T Y G G L P R K R V V E F F G
8867 cctgagccaagtgggaaactacttcagctctcgacattcgcaaatgagcgaatggtatttgagcaggaatgggaacagaaag
57 P E S S G K T T S A L D I V K N A Q M V P E Q E W E Q K
8951 actgaagaactcaaggaaaagctggaaaatgcgcgtgcatccaaagctagcaagactgctgtcaaggaacttgaatgcaactc
85 T E E L K E K L E N A R A S K A S K T A V K E L S M T R
9035 gatagcttcaagagcctcttaagattgtatatttgaccttgagaatacattagacactgagtggtggttaaaagattggagtc
113 D S L Q E P L K I V Y L D L E N T L D T E W A K K I G V
9119 gatgttgacaatatttgatagttcgccctgaaatgaacgcgctgaagaataacttcaatattgttttagacattttcgaaaca
141 D V D N I W I V R P E M N S A E S I L Q Y V L D I F E T
9203 ggtgaagttggcctagtagttctagattccttgcccttacattggtcagtcacaaacctatttgatgaagagttgactaaaaggcc
169 G E V G L V V L D S L P Y M V S Q N L I D E E L T K A A
9287 tatgcaggaatctcagcgctttagtgaatttagtcgaaggttactcctcttcttactcgctacaaatgcaatattcctagggc
197 Y A G I S A P L T E F S R K V T P L L T R Y N A I F L A
9371 atcaatcaaatcgagaagatagtaagtagtcagtaacatccctattcaactccaggcggaagatgtggaagactgctgttgca
225 I N Q I R E D M N S Q Y N A Y S T P G G K M W K H A C A
9455 gttcgacttaaattagaaaaggtgactaccttgacgaaaacgggtgcatcattgaccgctactgctcgaaacctgcagggaat
253 V R L K F R K G D Y L D E N G A S L T R T A R N P A G N
9539 gtagtagagtcattcgtcgagaagacaaagcatttaagccggacagaaaattagtttctctatcgcgttctcctatcatgagga
281 V V E S F V E K T K A P K P D R K L V S Y T L S Y H D G
9623 attcaaatgaaaatgaccttgtagatgtcgctgcaatttggagtcattcaaaaggcaggggcatggttcagatcgtcgac
309 I Q I E N D L V D V A V E P G V I Q K A G A W F S I V D
9707 cttgaaactggagaattatgacagatgaagcagaagaccattgaagttccaaggcaaggcaaatctagttcgacgcttcaag
337 L E T G E I M T D E D E P L K F Q G K A N L V R R F K
9791 gaggatgactacttattcgacatggtgatgactcggttcacgaaattatcactcgagaagaaggctaa 9859
365 E D D Y L F D M V M T A V H E I I T R E E G *

dp1ORF011

28017 atgaattttatgattatatacagcaggggagattgctagctacattcaagcacttcttcaaacgctcttcaaatccttggga
1 M N I Y D Y I N A G E I A S Y I Q A L P S N A L Q Y L G
28101 ccaactctttccctaatgctcaacaaacagggacagacatttcatggctcaagggtgcaataatttgcagtaactatccag
29 P T L F P N A Q Q T G T D I S W L K G A N N L P V T I Q
28185 ccattctaagcagcgcgaagcagcttctgtagcagctggtgatttagcaaacagctactgagatggcattcttccgtgag
57 P S N Y D A K A S L R E R A G F S K Q A T E M A P F R E
28269 tctatcgagcttggtagaaaagaccgtcaaaacttgcaaatgctattgaacaaagttcagctcttgcacaaaccttatcact
85 S M R L G E K D R Q N L Q M L L N Q S S A L A Q P L I T
28353 caactctataatgataactaagaacctttagacgggtgttgaagcgcaagcagaatacatgctgctgcaattgcttcaatcgggt
113 Q L Y N D T K N L V D G V E A Q A E Y M R M Q L L Q Y G
28437 aaattcactgtcaaatcaactaacagcgaggtcaatacacttacgactacaacatggatgctgaagcaacaatattcgactcact
141 K F T V K S T N S E A Q Y T Y D Y N M D A K Q Q Y A V T
28521 aagaatggactaaccagctgaagtgaccctatcgctgacatttttagcagcaatggatgacatcgaaaactgtacaggtgtt
169 K K W T N P A E S D P I A D I L A A M D D I E N R T G V
28605 cgccctactcgaatggcttgaaccgaaacacttataacaaatgactaagagtgactctatcaagaagctcttgcattgggt
197 R P T R M V L N R N T Y N Q M T K S D S I K K A L A I G
28689 gttcaaggttcttgggaaaacttcttcttcttcaagtgacgctgagaattcactcgctgaaaaaacaggtcttcaaatcgct
225 V Q G S W E N F L L L A S D A E K F I A E K T G L Q I A
28773 gtctacttcaagaaaattgctcagttcgctgacgtgacaaacttctgacgttggttaacattcgtcagttcaacttgatgac
253 V Y S K K I A Q F A D A D K L P D V G N I R Q F N L I D
28857 gacggtaagtggtattgcttccactcagcgagtggtgacacttggtagcgtactactccagaagcattcgacttggcttca
281 D G K V V L L P P D A V G H T W Y G T T P E A F D L A S
28941 ggcggaacagagcgtcaaggttcaagttcttccagcgagcctaccgttacaacttatcttgaaaaacctctgtcaacattgca
309 G G T D A Q V Q V L S G G P T V T T Y L E K H P V N I A

29025 acagttgtatcagctgttargattccatcattcgaaggaattgactatgtaggagttctcacactaattag 29096
337 T V V S A V M I P S F E G I D Y V G V L T T N *
dp1ORF012
5346 atgagttattaagttcaaaaccgaagaactttcaaaaattgtttctcagctcaataagttgaagcctagcaagttgctagaaatc
1 M S I K F K T E E L S K I V S Q L N K L K P S K L L E I
5430 acaaaactattggcatatttttggtgacggcgaatgcgtcatgtttacagcgtatgatggctcaaaacttcttcgatgcattatc
29 T N Y W H I F G D G E C V M F T A Y D G S N F L R C I I
5514 gacagcgatgttgaattgacgtgattgtgaagcagagcagtttggaaaactttagaaaaagaccacggcgcaaccgtcaca
57 D S D V E I D V I V K A E Q F G K L V E K T T A A T V T
5598 ttagtctcgaagaatcttcgctaaaagttattgggaatggtagtacaatattgatattgttacagaagatgaagagtagccct
85 L V P E E S S L K V I G N G E Y N I D I V T E D E E Y P
5682 acattcgaccacttgcctgaagacgtgagtgaaagaaatgcctcactttgaaaagctcgctgttctacggaatcgccaatctc
113 T F D H L L E D V S E E N A L T L K S S L F Y G I A N I
5766 aacgattctgcggtatctaaatcaggagcagatggaatttataccggcttctgtttaaaggcggaagcaattactacagac
141 N D S A V S K S G A D G I Y T G F L L K G G K A I T T D
5850 atcattcgcgtatgtatcaaccctatcaaggaaaaggtagaataatgctcattccttacaacctaatgagatttttagcaagt
169 I I R V C I N P I K E K G L E M L I P Y N L M S I L A S
5934 attcctgatgagaagatgtactctggcaaatgtacgatactactgtctatatttcacgtgcttcagtcgaattttatggaaaa
197 I P D E K M Y P W Q I D D T T V Y I S S A S V R I Y G K I
6018 ttgatggaaggtatggaagattatgaagacgtttcacagcttgactcaattgagtttgaagatgatcggttatccctacagca
225 L M E G M E D Y E D V S Q L D S I E F E D D A A I P T A
6102 gaaattcctgagcgtatttagaccgcttgccttacttactcacttcagcctttgacaaaggaaacgtcgaaattcttattctgaaagac
253 E I L S V L D R L V L F T S A F D K G T V E F L P L K D
6186 cgacttcgaattaaaacttctactagcagttatgaagacatcatgtacgcatctgctggcaagaagtttcgaagaaagaaatc
281 R L R I K T S T S S Y E D I M Y A S A G K K V S K A E F
6270 acttgccaccttaacagcttactctgaaggaattgtatcaaccgtcaccgaagaaaacttactgtctcttatggaagcgaa
309 T C H L N S L L K E I V S T V T E E N F T V S Y G S E
6354 accgcaattaaagatttcacgaatgggtgcgttacttctcctagcacttcaagagccggaagataa 6419
337 T A I K I S S N G V V Y F L A L Q E P E E *
dp1ORF013
10215 atgaatttagcttctaaataccgtcctcctcaaaactttcgagggaagtgtagctcaagaatatgtcaagaattcttttgaatcaa
1 M N L A S K Y R P Q T F E E V V A Q E Y V K E I L L N Q
10299 ttacaaaatggcgctatcaaacacggctatctattctgtggtggcgctggaactggtaaaaccactactgctcgaattttcg
29 L Q N G A I K H G Y L F C G G A G T G K T T T A
10383 aaggatgtgaacaaaggacttggctctcctattgaaattgatgctgcttctaataatgggttagaanaatgttcgaacattatt
57 K D V N K G L G S P I E I D A A S N N G V E N V R N I I
10467 gaagattcctagatcaagtcctatggacagcgagttcaaaagtttcatcattgacgaggttcatatttcaaccgggaatcatt
85 E D S R Y K S M D S E F K V Y I I D E V H M L S T G A F
10551 aatgcgctgttgaanaacattagaagagccctcatcggaacccgtgttctattctatgtactactgaccctcaaaagattcctgac
113 N A L L K T L E E P S S G T V F I L C T T D P K I P D
10635 actattctcagtcgagttcaacgggttgactttactcgaattgataatgaacacatcgttaacacttcaattttatctgaa
141 T I L S R V Q R F D F T R I D N D D I V N Q L Q P I I E
10719 agtgaaaatgaagaagagctgttattatgattgacgctgacgcttctgcttatttgggaacttgcaaatggaggaatgcgt
169 S E N E E G A G Y S Y E R D A L S F I G K L A N G G M R
10803 gacagttatcacaaggctcgaaaaagtccttgattatagtcacatgcgttgacatggaagccgttctaatgactaggagttccg
197 D S I T R L E K V L D Y S H H V D M E A V S N C A G V P
10887 gactacgaacattcgtcttacttggtaagctattgccaactatgacggctcaaagtggtttagaanaatgtaaatgacttccac
225 D Y E T F A S L V E A I A N Y D G S K C L E I V N D F H
10971 tactcaggaanaagacttgaatttagtgactcgaactttacagacttctcttttagaggtttgttaagtattggctagttcgagat
253 Y S G K D L K L V T R N F T D F L L E V C K Y W L V R D
11055 atttcaatcactcaacttctgctcttttgaagtaagctagagcaattctgtaggcttttcaatctcacttacttctggtg
281 I S I T Q L P A H F E S K L E Q F C E A F Q Y P T L L W
11139 atgctagaagaataatgaacttgcgtgagttgttaaatgggagcctaatgctaaaccgataattgaaaccaaacttcttttg
309 M L E E M N E L A G V V K W E P N A K P I I E T K L L L
11223 atgagcaaggaggagtgga 11240
337 M S K E E *
dp1ORF014
50961 atgaagtaaatggtcttcaaatgaagcagactcctgaacaaataattgaaaaactttcgagacaacttgaagcgaaggaaca
1 M K V N G L Q I E A T P E Q I I E K L S R Q L E D E G T
51045 ttcatttttagcgaactaagtcgcttggaaagcaactcaattctcatgcccgtttcatgagggaggtgaaagcatccc
29 F I F R R T K S L G S N Y O P S C P F H A G G T E K H P
51129 tcttggtgagtagtagaaatccttcttattcaggaagtaaggtgacggaagctggaacggttctactgttctacttgcggctac
57 S C G M S R N P S Y S G S K V T E A G T V H C F T C G Y
51213 acttcaggactaactgaattcgtctcgaatgtattaggtcgaaacgatggagggttctatggaaacagtggttgaaaaggaat
85 T S G L T E F V S N V L G R N D G G F Y G N Q W L K R N
51297 ttggaaacatctagcgaagtagttaggcaagcgctcagccctgaagcgttttcgaagaataagggaactgaaagtagtcgagcat
113 F G T S S E V V R Q G V S P E A F R R N G R T E K V E H
51381 aaaactcattcctgaagaggaacttgataaaatccggtttattcattccttatatgtatgaacggaaattgacggagcagctc
141 K I I P E E E L D K Y R F I H P Y M Y E R K L L D E - 4 I
51465 gagatgtttgatgtaggttatgacaaactgcagtgatgcacctttccagtagcgaacctcaaggcggaacagtagttcttc
169 E M F D V G Y D K L H D C I T F P V R N L K G E T V F F
51549 aaccgtcgaaagtgcttctaaagttccaccagtcggtgaaagtagaccctaaaacgggaatttcttctggccaatgagctt
197 N R R S V R S K F H Q Y G E D D P K T E F L Y G Q Y E L
51633 gtgacatttcgagactattttgaaaaacttattgtaagtagtgcgtgagctgagttcttatcaactgcttacttcttgggtca
225 V A F R D Y F E K P I S Q V F V T E S V I N C L T L W S
51717 atgaagattccagcagtcgctcttatgggagtaggtcggaagaaatcaaatcaatttactaaaacgacttcttatagaataatt

371

253 M K I P A V A L M G V G G G N Q I N L L K R L P Y R N I
51801 gttctagcacttgaccctgataacgctgggcagacagcgaggaactctaccgacagctaaagcgaagcaggtcggtaga
281 V L A L D P D N A G Q T A Q E K L Y R Q L K R S A V V R

51885 tttttgaactaccctaaagagttctatgataataagtgggatataaacgaccatccggaattattaaattttatgatttagtc
309 F L N Y P K E F Y D N K W D I N D H P E L L N F N D L V
51969 ttgtag 51974
337 L *

dp1ORF015
1793 atgggatttaattctatacttcgcaggaggtcacgctatttagcactgacgattatttgaaggaaagaggacccatcgctatttc
1 M G F N L Y F A G G H A I S T D D Y L K E R G A N R L F
3877 aatcaactgtacgaagaaacgggattggcagaagtggtgagcataagaaaaccaatccaagcactacttcaaaactatttc
29 N Q L Y E R N G I G K R W I E H K K T N P S T T S K L F
3961 gtcgactctagtgcatattctgctcataccacaaaggggctgaagttgacattgacgcttatatcgaaatcagtgataacgtg
57 V D S S A Y S A H T K G A E V D I D A Y I E Y V N D N V
4045 ggaatgtttgactgtatcgccgaactcgataaaattcctgggtgatttagacagcctaagacacgtgaacagcttttggaaagca
85 G M F D C I A E L D K I P G V F R Q P K T R E Q L L E A
4129 ccacaaatttctgggataattatctatcacatcgccgagcgaatgggtgagaaagacaagctcttacctattttccatattggga
113 P Q I S W D N Y L Y M R E R M V E K D K L L P I F H M G
4213 gaagactttcaatggctcaacttgatgctcgaaactacattcgaaaggcgaagcatattccttacattggaatttcaccagcc
141 E D F K W L N L M L E T T F E G G K H I P Y I G I S P A
4297 aatgactcgactacgaagcatcaagacaagtggtggaagagatttcgaagttattcgaaacagttcctaaccagacgttaag
169 N D S T T K H K D K W M E R V F E V I R N S S N P D V K
4381 actcagcatttgggagtgacagttactagccaattagagcgctacccattctatagcgccgactctactctgtactgctcaca
197 T H A F G M T V T S Q L E R H P F Y S A D S T S V L L T
4465 ggagcagtggaagcaattatgacgtcaaaaggattagttgacttgcacagaagaatggaggaattgatgctgctcgtaggctg
225 G A M G N I M T S K G L V D L S Q K N G G I D A V R R L
4549 ccaaaaccggttcaagttgaaattgaatccattatcgaaagaaactggagcgcattttagcctagagcaatttagtgaggactat
253 P K P V Q V E I E S I I E E T G A H P S L E Q L V E D Y
4633 aaacttcgagcattgttcaatgttcaatacatgctgaattgggcagagaactatgaattcaagggaattaaaaatcgctcaacgt
281 K L R A L F N V Q Y M L N W A E N Y E F K G I K N R Q R
4717 cgactattttag 4728
309 R L F *

dp1ORF016
43413 atgggagtcgatattgaaaaaggcgttgctggatgcaggcccgaaagggtcgagttatcttatagcatggactttcgagacggt
1 M G V D I E K G V A W M Q A R K G R V S Y S M D P R D G
43497 cctgatagctatgactgctcaagttctatgtactatgctctccgctcagccggagcttcaagtgtggtgggagctcaactact
29 P D S Y D C S S S M Y Y A L R S A G A S S A G W A V N T
43581 gagtactgcacgcatggcttattgaaaacgggttatgaactaattagtgaaatgctccgtgggagctcaaacgagggcagatc
57 E Y M H A W L I E N G Y E L I S E N A P W D A K R G D I
43665 ttcatctggggacgcaaaagggtgctagcgcaggcgtggaggtcatcacagggatgttcattgacagtgataacatcttactgc
85 F I W G R K G A S A G A G H T G M F I D S D N I I H C
43749 aactacgctacgacggaatttccgtcaacgacacgagtgagcgttggtactatgcaggtcaaccttactactacgtctatcgc
113 N Y A Y D G I S V N D H D E R W Y Y A G C Q P Y Y V Y R
43833 ttgactaacgcaaatgctcaacggctgagaaagaaacttggctggcagaagatgctggtttctggtacgctcgagcaaac
141 L T N A N A Q P A E K K L G W Q K D A T G P W Y A R A N
43917 ggaacttatccaaaagatgagttcgagttatcgaaagaaacagtccttggttctactttgacgaccacaggtcacatgcgct
169 G T Y P K D E F E Y I E E N K S W F Y F D D Q G Y M L A
44001 gagaaatggttgaacatactgatgaaatttggtatttggttcgacccgtgacggatcatggtcagctcaggaacggatggc
197 E K W L K H T D G N W Y W F D R D G Y M A T S W K R I G
44085 gagtcattggtactacttcaatcgcgatggttcaatggtaacgggttggaattagttacgataattggattattgtgctgct
225 E S W Y Y F N R D G S M V T G W I K Y Y D N W Y Y C D A
44169 accaacggcgacatgaaatcgatgcttataacgacggctggtatctactattaccggacggacgtctgagcat
253 T N G D M K S N A P I R Y N D G W Y L L L P D G R L A D
44253 aaacctcaattcaccgtagacggcggcggtcattactgctaaagttaa 44303
281 K P Q P T V B P D G L I T A K V *

dp1ORF017
11242 atgattggacagggacttgtaaatctaccatttcgaaatggaaacaaacttccaaatatataatcgtcgaaggcgaagtaggt
1 M I G Q G L V K S T I S K W K Q L P K Y I I V E G E V G
11326 tcaggacggaagacctaactccgttatattgcttcgaaatttgacgctgattctattgtagtaggaacgagtgtagacatt
29 S G R K T L I R Y I A S K P D A D S I V V G T S V D D I
11410 cgaacatcattcaggatgcacagactattttcaaggcgagaatctacgtgatagcggaaatagcctgtcaatgtcagctctt
57 R N I I Q D A Q T I F K A R I Y V I D G N S L S M S A L
11494 aactcgcttttgaagatagcgaagagccacctttaaactgtcatatagccatgactgttgatagcatcaaatgctttacct
85 N S L L K I A E E P P L N C H I A M T V D S I N N A L P
11578 acgcttgcaagtagagcaaaagttctaaccatgctaccttataactaagaaagaaaatgcagtttgtcaagctctacaagaag
113 T L A S R A K V L T M L P Y T N E E K M Q P V Y K K
11662 gtatgacttccaggaattgacgacccgagcgttgtagactattgcaatcttgccagcaatcttcaaatgcttgaagacatatta
141 V D T S G I D D R A I T G D Y C N L A S N L Q M - L E - J L
11746 gaatatggcgagaagagctatttgaagaggttacacattttatgacttaatatgggaggcaagtgctaggaaattcgctaaag
169 E Y G A E E L F E K V T T F Y D L I W E A S A S N S L K
11830 gttactaattggctcaaatttaaggaaactgatgaaggaataattgagcctaacttttctcaactgcttttaaatggctg
197 V T N W L K F K E T D E G K I E P K L P L N C L L N W S
11914 acagttgtcatcaggaagcactatgtagaatgtcttcgaagaacttgaggcccatgaccttttagtgagggaagcattcagg
225 T V V I R K H Y V E M S F E E L E A H D L L V R E A S R

11998 tgtttgcgaaggatctctaaaaagggtcaaatgcccgtgctgctgacgaacgaatttatcaggagggtcaacaagttgagtga
12081
253 C L R K V S K K G S N A R V C V N E F I R R V K Q V E *
dp1ORF018
35847 atggctagcagacagacgctatttggtcgacggaattgaccttgcgacaaagggtgcaaccgtgctagaatatgtaggactcact
1 M A S R Q T L L V D G I D L V D K G A T V L E Y V G L T
35931 ttcgcaggatttaaggactcaggatttataaaaccctgaaggcatagacggagtagtagatttccgctcctaatgctatgctccgct
29 F A G P K D S G F K N P E G I D G V L D S P S N A M S A
36015 cttactggaagcgtgaccttaattgtccacgggagaaaccgaaagcaagttaatcaaaaacacaggcagttcaacaatttatt
57 L T G C S V T L M F H G E T E K Q V N Q K Y R Q P K Q F I
36099 cgctcgaagtcattttggagaatttcgacacttgaagacctggatactatcgaacgggaaaaatttttaggagaaaacgagcaa
85 R S K S F W R I S T L E D P G Y Y R T G K F L G E T E Q
36183 ggaaaactttagacgttcaagcctttaaagatacttccctttagttaaattagggttagttaaagatgcttacgagtac
113 G K L V D V Q A F K D T S L V V K L G I Q P K D A Y E Y
36267 agcactcaactgttcgaaagggttataagtttcaacccgcttggaggcgatagcttacctaacccggaagacctactcga
141 S D S T V R K V Y K F Q P A L G G D S L P N P G P T R
36351 caatttagagtagaataagaactacttctcaaatcaaggatattttcgaattggcgaaaaaagttcaggacagtttgttgag
169 Q F R V E I R T T S Q I K G Y P R I G E K S S G Q F V E
36435 ttcggtacaaatcagtagttagtgaaggtggcgtgatttatttcaaatcttgaacttttgaactttataaaatcagtagt
197 F G T N S V L M E S G S I I I L N L G T F E L I K I S S
36519 gcaaatcaagcgactaacttatttagatataaaccgagggcgatttctcaagattcctaattggaattcaacaattaccatt
225 A N Q A T N L F R Y I K R G A F P K I P N G N S T I T I
36603 gaataccgagccgatgacgcagcagcttggacctctactcttccgctcaagttgaactgtttctaaatccgtcttactattag
36686
253 E Y R A D D A A A W T S T L P A Q V E L F L N P S Y Y *
dp1ORF019
12161 atgaatgtttatctcaatcaaatgggaatgtagttcgagaaacttcgggtttcaacagctcggaaaaaccctcactcaaaaaggg
1 M N V Y L N Q M G N V V R E T S V S T V W K T L T Q K G
12245 ctctgttctaatcatcgaatattcgtgcttcgagatgataaggagtttctgtctaatgagtcgaggtggaaaaagggttccggat
29 L V S N H R I F A V R D D K E F L S N E S R W K R L P D
12329 gtttagatgggacacttgtttgattggttactaaaattgacaagcgaagtgcttaaggccctttctgataatttctgtt
57 V R Y G T L V L M V T K I D K R S K L L K A F P D N C V
12413 gagtttgagaaaatgactgacgcagttgaaaaggcattttgtgtctaaatctcgaactattgatagcgacatgactgacatg
85 E F E K M T D A Q L K R H F V S K Y S T I D S M I D M
12497 gttatccagttctgtctaaacgattactctagaattgacaatgaattggacaagctgtcgcgattgaaaaagggttgacgcacatca
113 V I Q P C L N D Y S R I D N E L D K L S R L K K V D A S
12581 gttagttgaatccattgtcgaagcacaagaccgaaattgacattttcagccctagttgatgatttggaaattataggccggagcag
141 V V E S I V K H K T E I D I F S L V D D V L E Y R P E Q
12665 gcaatttgaaagtgactgaacttttagccaaaggaggaagtcctattggattgcttacctgttcttatacaataaac
169 A I M K V T B L L A K G E S P I G L L T L L Y Q N F N N
12749 gcttctctgtgtaggagccgatgagcctaaagaagccaatctaggcattaaagcagttcttaatacaataagattgcttataac
197 A C L V L G A D E P K E A N L G I K Q P L I N K I V Y N
12833 tttcaatacagagctggactcagcctttgaaggcagtgctatttttaggtcaagctatcgagggcataaagaatggtcgctataca
225 P Q Y E L D S A P E G M A I L G Q A I E G I K N G R Y T
12917 gaaagtccagtggtctatatttcttctataaaaatttttctactactaa 12967
253 E S S V V Y I S S L Y K I F S L T *
dp1ORF020
1864 atggctaatcaatacaatcagcctgaaaggaggaagattcgaatcaatgttcgcgaccctgagaaaaatgcctatcatggaatt
1 M V N Q Y N Q P E R G K I R I N V R D P E K M P I M S I
1948 ttcggtctcaaatcaagggtgaaggaaattggtatcgaagactattttcattcgaactgggtgagtagcagactattgct
29 F G P T I Q G E G M V I G Q K T I F I R T G G C D Y H C
2032 aactggtgtagcagcctttacctggaacggtagtactgagccggaatatatcacaggcaagaagctgtagtgaatcttg
57 N W C D S A F T W N G T T E P E Y I T G K E A A S R I L
2116 aaactagctttcaatgataaagggtgaacagatttgaaccacgtgacattgactggaggaatcctgccttaatacaacgagcct
85 K L A F N D K G E Q I C N H V T L T G G N P A L I N E P
2200 atggctaaagatgatttcgattctaaaagaacatggattcaagtttggtctcgaaaactcaaggaaactcgattccaagaatggttc
113 M A K M I S I L K E H G F K F G L E T Q G T R F Q B W F
2284 aaagaagtaagcgatatcactattagtcctaaaccgcttcaagtggaaatgagaactaatatgaaatcttgaagctattgta
141 K E V S D I T I S P K P P S S G M R T N M K I L E A I V
2368 gatagaatgaatgatgaaaaccttgactggtcattttaaactcgttatctttgacgaaaatgacctagcttatgctgctgatatg
169 D R M N D E N L D W S F K I V I F D E N D L A Y A R D M
2452 tttaaaactttcgaaggcaagttacgtccagtgaaactacctttcagttgggaatgcaaacgcatacgaagaaggaaaatcag
197 F K T F E G K L R P V N Y L S V G N A N A Y E E G K I S
2536 gataggcttcttgaagggtggatggctttgggataaagtgtatgaagaccagctttcaacaatgttcgacctttaccgcaa
225 D R L L E K L G W L W D K V Y E D P A F N N V R P L P Q
2620 cttcatcacacttgtttatgataataaaggaggtataa 2658
253 L H T L V Y D N K R G V *
dp1ORF021
2504 atgcaaacgcatacgaagaaggaaaaatcagtgataggcttcttgaagggtggatggcttgggataaagtgatgaagacc
1 M Q T H T K K E K S V I G F L K S W D G F G I K C M K T
2588 cagcttttcaacaatgttcgacctttaccgcaacttcatacacttgttatgataataaaggaggagataaaatgaaattgag
29 Q L S T M F D L Y R N F I H L F M I I K E E Y K M K I E
2672 catctagataaaaatcggtaacgtatttagggagagagacggatggcttcccttaagccggatgaaattgtaaccttggacaat
57 H L D K I G N V L G R E N G W A S L K P D E I V T P D N
2756 actgaggcagccgttcaagacttttgggtctattaggcgaggacgagaacgtgacgggttgaagatactccattccgtttt
85 T E A A V Q R L F G L L G E D A E R D G L Q D T P F R F

2840 gttaaagcactcgctgaacataccgtagggtatcgagaagaccctaaacttcatctcgaaaaacattcgacgtcgaccatgaa
113 V K A L A E H T V G Y R E D P K L H L E K T F D V D H E
2924 gacctgttctgtgaaagacattccattcaattctttatgtgagcatcatttagctccgttcgtagggaaggtgcatattgca
141 D L V L V K D I P F N S L C E H H L A P F V G K V H I A
3008 tacattccctaaggataagattacaggtctttcaaaattcggtcgagtggtggaaggatacgttaaacgacttcaagtacaagag
169 Y I P K D K I T G L S K F G R V V E G Y A K R L Q V Q E
3092 cgcttgactcaacaaatcgctgacgctattcaggaagtcttaaatcctcaagcagttcggtcatcgtagaggctgagcact
197 R L T Q Q I A D A I Q E V L N P Q A V A V I V E A E H T
3176 tgcagcgacgagcggttataaagcacggggcaacgacagtgacttcaactatcgaggtcttttccaagatgacgcatct
225 C M S G R G I K K H G A T T V T S T M R G L F Q D D A S
3260 gctcgagcagaattgcttcagttgattaaaaagtag 3295
253 A R A E L L Q L I K K *

dp1ORF022
30896 atgagtaaaagacattctttacggaatcaagctcggtgcaaatcgaggagcttgacctgactcagttgctcaaaagtcggcgga
1 M S K D I L Y G I K L V Q I E E L D P L T Q L P K V G G
30980 gctaactttgtcgtagatacggcagaaacagcagaactcgaagccgtgacctcgagggaactgaagatgtgaaacgcaatgac
29 A N F V V D T A E T A E L E A V T S E G T E D V K R N D
31064 acgctgactccttgcctatcggtcggtacgacacgttttatacgggttatgacttaacattcaaggacaacagttgacctgaa
57 T R I L A I V R T P D L L Y G Y D L T F K D N T F D P E
31148 atcatggccctaatgaaggtggtacagtcagtaacaaggcggaactattgctggatagacaccccaatgcttgacacaggt
85 I M A L I E G G T V R Q Q G G T I A G Y D T P M L A Q G
31232 gcttctaattatgaaacatttgaatgaacatcttggtgcaaaactatgtaggtgactcaattgtcaactacgtgaaatcact
113 A S N M K P F R M N I Y V P N Y V G D S I V N Y V K I T
31316 tgaataaactgtaccggttaaagctccagggtcttcaatcggaagagttctacgctcctgagttcaacatcaaggcacgtgaa
141 L N N C T G K A P G L S I G K E P Y A P E F N I K A R E
31400 gcaaccaaagcaggtttgacagttcaagtcaatggactatgtggcacaaactccagcggtcttctgctcgctgacattcgattg
169 A T K A G L P V K S M D Y V A Q L P A V L R R V T F D L
31484 aacggtggaacaggaacccgacgacgttccagtgtaagcaggttaagaagatttctccaaacaggttgacctctacttaaca
197 N G G T G T A D A V R V E A G K K I S P K P V D T P L T
31568 ggtaaggctttcaaggctggaaggtgaaggagaatcaactatttgggacttcgacaaccacatgatgcctgacggagacgtc
225 G K A F K G W K V E G E S T I W D F D N H M M P D R D V
31652 aaactcgtagcacaatttgcatag 31675
253 K L V A Q P A *

dp1ORF023
6419 atggccaagtcgaatttaactagaattgcaagatgggttagagcaggaacagtgagggtcctgcttcatcttttgcattctg
1 M A K S N L T R I A K M V R A G N S E G P A S S F V N S
6503 ctgacccgggttattgaacgaactcagctcgaatataatccttcgacatattataagcccgaggggttggtggtgatttga
29 L T R V I E R T Q P E Y N P S T Y Y K P S G V G G C I R
6587 aaaatgtatttgcgaagaatcggtgagttctattatagataacgcagattctaactaatgcaatggcggaagctggaacatt
57 K M Y F E R I G E S I I D N A D S N L I A M G E A G T F
6671 aggcacgaagttctcgaaggtacatgggttaaaatggctgaaatcgatgaggactttgaatgggtgaatgtagcagagttcttg
85 R H E V L Q E Y M V K M A S I D E D F E W L N V A E F L
6755 aaagaaaatccagttgaaggaactatcgtcgacgagcggttcaagaaaaacgattatgaaacgaagtgtaagaacgaactctt
113 K E N P V E G T I V D E R F K K N D Y E T K C K Q V L L
6839 caactttcattcttgtgacggactagttcgatataaaggcaagctctacatttttagagattaaagactgaaacattgctcaag
141 Q L S F L C D G L V R Y K G K L Y I L E I K T E T M F K
6923 tctgctaacaactagcagcctatgaagaacacagaagtgcaagcaacttgctacggaatgtgtctaggagtagcagctcatt
169 F T K H T E P Y E E H K M Q A T C Y G M C L G V D D V I
7007 ttcctttatgaaatcgagataaacttcgaaaagaagcctacacgtttcacatcacagacgagatgaaaaatcaagtccttgga
197 F L Y E N R D N F E K K A Y T F H I T D E M K N Q V L G
7091 aaaattatgacctgcgaagagtagtagagaagggcgaaggtcctaaaatctattgctcttcagcctattgcccattttaga
225 K I M T C E E Y V E K G E S P K I Y C S S A Y C P Y C R
7175 aagggaaggtcgaaatctgtga 7195
253 K E G R N L *

dp1ORF024
25992 atgaacgcagtagatggccaggtagttcatattctacaagatttagcagaagatggaaatgctacgggtgaaaagttcgaaaag
1 M N A V D G Q V V H I L Q V L A E D G N A T A E K F E K
26076 gaagtcagggtcgcatcttttagtattttcacgaagagcagccgaggttgctcaaggtgaaatctataaggacggcgaacaaac
29 E V R A A S L V F S R R A A E A V V K G E I Y K D G K N
26160 ctctcgaaacgtgttgggtcttcagccgcagcgcaagaaatgatgttcaacaaatagtcacacaaggcctagcaggaatg
57 L S K R V W S S A A R A G N D V Q Q I V T Q G L A S G M
26244 tctgctacagatattggctaaaatgctcgagaatatatcgaccctaaaggttcgaaaagattgggactttgataagatagctgag
85 S A T D M A K M L E K Y I D P K V R K D W D F D K I A E
26328 aagctagggaaaacctgctgctcataaatatcaaaatctcgaatacaatgcccttcgacttgctcgaactaccatttagccattcc
113 K L G K P A A H K Y Q N L E Y N A L R L A R T T I S H S
26412 gccacagctggaagtgagacaatggggcaaggttaatccttatgctcgaaaagttcaatggcattctgttcacgctccaggtcga
141 A T A G V R Q W G K V N P Y A R K V Q W H S V H A P G R
26496 acgtgtcgaagcgtgtatcgatttagatgggtgaagttatcttatcgaagaatgtccttcgacctcctaatggaatgtgtac
169 T C Q A C I D L D G E V F P I E E C P F D H P N G M C Y
26580 caactgtatggtacgaaaactcactcgaagaatcgctgaggttgagaggtgggttagacggagaacctaactgatttatta
197 Q T V W Y E N S L E E I A D E L R G W V D G E P N D V L
26664 gacgaatggtacgacgatttaagttcaggaaaagttgagaaatacagcgacctcgacttgtttaaagttattag 26738
225 D E W Y D D L S S G K V E K Y S D L D F V K S Y *

dp1ORF025
18778 atggcaagaacaaaaagcgaaaaaagtaaatgtcaaaaggaaatgcttatccctacaaatctctcgaaaaaagtaaatgta
1 M A K N K K R K K V N V K R K M L I P T N L S K K V N V

18694 aaagcaatcgcttatagaaaagtcactgttaagtggtgcctaatacagatgaaattcaagtatatttcgacctttatataaat
29 K A I A Y R K V T V K W L P N T D E I Q V Y F D L Y I N
18610 aaaaacaggctgacaatgttaggcactattgacccggacaaagagctattttgaaggaattaggattgtttgtaagaaacccag
57 K N R L T M L G T I D P D K S Y F E G I R I V C K K P Q
18526 ccttggaatgactgttaagagctccagggttcgctgcagacgccccaggttttttgcagttcttaagcctattgtcacacg
85 P W M T V K E L Q V A R A D A P G F F A V L K A Y C H T
18442 gttggcagtgactagatagcggagcagagcctactgaaattgttcaaggtattatgtataaagacgggtgaactatttaaggac
113 V G D V L D S G A E P T E I V Q G I M Y K D G E L F K A D
18358 agtgaattgtcagccttttcaaatcagatgtcaagagccttatgagtttccaaaggaccttccatataaccttggacaacttt
141 S E I V S L F K Y D V K E P Y E F P K D L P I T L D N F
18274 tttagagttcattatgtctagccagcactagagcactgttttgcgtttgtgtaataataggtgagttttccaagaattggcgg
169 L E F I M S S Q H T R A L V L R C A N I G E F S K N W R
18190 aaatggcaaaagctatccagctcctgctcagactgtccaaggcggatgactttaaagtagacgaaactgtttgggacttttca
197 K W Q K A I Q L L L D Y A K A D D F K V D E T V W D F S
18106 cccggctctaaagctggaaaggtagcagctcgttaaaggctatgaggcaattcaacaagcccttgagcagataaaataataa
18026
225 P G S K A G K V A R R K G Y E A I Q Q A L E Q I N K *
dp1ORF026
21512 atggcgaaagctactggaccaaaagttcgaagaggaaaaactcctccacggccaaaagacaaaaaggaatcaaaagcaaatgcg
1 M A K A T G P K V R R G K T P P R P K D K K G I K A N A
21596 cgtgtcaataaagaccagttcgttagagtattgactaacaggcatcaagatgacaattaaagaaacgtgattgctagaatgaaattg
29 R V N K D Q F V E Y D Y K G I K M T I K E R D A R M K L
21680 gaattttattagagcactgactattcaggaaattgcagcccgctatggattaaatgaaaagcgtgttggcgaaatcagggtcgc
57 E F I R G M T I Q E I A A R Y G L N E K R V G E I R A R
21764 gataaatgggtgaaggctaaagaaagagttcagaaatgaaaagcctcgttactaatgatacattgactcaaatgtatggcagg
85 D K W V K A K K E P E N E K A L V T N D T L T Q M Y A G
21848 tttaaagctcagtcattataatatacagcccgctgggagaaactaatgaacatcgtcgaaatgtgttagataatcctgac
113 P K V S V N I K Y H A A W E K L M N I V E M C L D
21932 agatattttatttactaaagaaggaaatattagatggggcgcttagatgtcctttcgaaaccttatagatagagctcaaaaagga
141 R Y L F T K E G N I R W G A L D V L S N L I D R A Q K G
22016 caagaaagagcgaatggaatgcttccggaagaggttcgataatagactacaaattgagcgcgagaaatgacattgtccggcc
169 Q E R A N G M L P E E V R Y R L Q I E R E K I T L L R A
22100 aaaaatggcgacccaggaattgaaggcgaggttaaagataaacttcgtagaagcactagataaagcagctcaagcgtttggcaa
197 K M G D Q E I E G E V K D N F V E A L D K A A V W Q
22184 gaatttagtgacgcaacaggttctcatataaaggagtgactgataatgacaataagcctgagaaataa 22252
225 E F S D A T G S Y I K G V T D N D N K P E K *
dp1ORF027
52762 atgggaaaaagttatcaattcaaaaatcaggaacatttagctcaggggtctaataacagagtttttcacactcgctgaccacgggtgac
1 M G K V S I Q K S G T F S S G S N N E F F T L A D H G D
52846 agcgcgaattgtcactctattgtatgacccggaagacatggattatttcgttagtccacgaagcagacgttgacggt
29 S A I V T L L Y D D P E G E D M D Y F V V H E A D V D G
52930 cgtcgacgcttatcaattgcaatgctattggcgaaagcggggaacagtcctcctgataattgtccattatgcccacggga
57 R R Y I N C N A I G E D G E T V H P D N C P L C Q N G
53014 ttccctcgtattgaaaaactatttctcaactttcaaccatgatacgggaaaagttgaaacatgggaccgagcgttcttat
85 F P R I E K L F L Q L Y N H D T G K V E T W D R G R S Y
53098 gttcaaaagattgttacatttatcaataaatatggaagccttgtgactcagccttttgaaattatttcgttcaggagctaaagg
113 V Q K I V T F I N K Y G S L V T - Q P F E I I R S G A K G
53182 gaccacgaactactatgaattccttcagagcgtccggaagacagtgctactcttgaagattttccagaaaagagcgaactt
141 D Q R T T Y E F L P E R P E D S A T L E D F P E K S E L
53266 cttggaactctaattttagacctcgacgaagaccaaattgttgacgtggttgacgggaagttcacttccaagaagagcgttct
169 L G T L I L D L D E D Q M F D V V D G K F T L Q E E R S
53350 tcaagtcgttcaaatcagctagaggagcactctcgcgctagacgaggttcgggtcgagaatcttcacaaggtcgaacagct
197 S S R S N S R R G A S P A P R R G S G R E S S Q G R T A
53434 gaaagaactccttcagttagtcgaagaactcctccaacacgaggtcgaggtattctaa 53490
225 E R T P S V S R R T P P T R G R G F *
dp1ORF028
44595 atgtcaaaaattaaattcgaaaaaccttaaaaaaggcgatgttgtgctacgagctaaatctcaaacgaagtttaaaatcgtttca
1 M S K I K F E N L K K G D V V L R A K S Q T K F K I V S
44679 attttagcagacgaaaagaaagcagaccttgaaatcattagaagacggaggtgaactttcacctttcagcttcaactctcgaacgt
29 I L A D E K K A D L E S L E D G G E L H L S A S T L E R
44763 tggtagacaatggaagatgaaactgaacctaaaaaagaaagagctgctaaacctgctaaaaagggtgctcctgagttgctcga
57 W Y T M E D E T E P K K E E A A K P A K K A A P A V A R
44847 cctgctcgaaaaggtagagtcgttcccaaacctaaaaaagaagtccttgaggagaagaattcctgaagttaaaggaacagccggaa
85 P A R K G R V V P K P K K E V L E E E I P E V K E Q P E
44931 gaagttggttcagttagtgagaaatcactgttcgcaaaacctgctcctaaaaaagaagcgtgagttcgtgactaagagcttct
113 E V G S V S E K S T V R K P A P K K E S V M A I T K A L
45015 gaaagtcgaattgttgaagcctttctcgtctactcgaactcgtcactcagttacatcgctatcgctctaagaagaacttc
141 E S R I V E A F P A S T R I V T Q S Y I A Y R S K N F
45099 gttactatcgaagaactcgaagaagggtgtttctattggagttcgcgcaaaaggggtgacagaagaccnaaagaacacbtcttgca
169 V T I E E T R K G V S I G V R A K G L T E D Q K K L L A
45183 tctattgctcctgcatcttacgaatggcgaggtttttaaactcgtcaaggagaagatattgacacccgaatggaa
197 S I A P A S Y E W A I D G I F K L V K E E D I D T A M E
45267 ttgattgaagcttctcacctttcttcgctatga 45299
225 L I E A S H L S S L *
dp1ORF029

662 atgaaatcagtagtctttattatccggcgagtcgactcagccactcggttttagcaattgaagctgcacaagtggggtctcaaaat

1 M K S V V L L S G G V D S A T C L A I E V D K W G S K N

746 gttcactgcctagatcctaattcacggacaaaagcatgaagcagaacttgaaaatgctgcttaagtgttgcgaatgtctcagggagtc

29 V H A I A F N Y G Q K H E A E L E N A A N V A M F Y G V

830 aaagtccaccattctgaaattgcatcgaaaactctactcaagctctagctctctcttattacaaggaaggcgaaattccatcat

57 K F T I L E I D S K I Y S S S S S L L Q G K G E I S H

914 ggaaaatcttcacgctgaaattcttagcagagaagggaagtagttgacacctatgttccatttagaaaatggagctaatgtctcacag

85 G K S Y A E I L A E K E V V D T Y V P F R N G L M L S Q

998 gctcgggcttatgcttattcgggtggagcttcttactcgctgattatgggtgctcacgcagacgatcgggctggagggtgcttacctt

113 A A A Y A Y S V G A S Y V V Y G A H A D D A A G G A Y P

1082 gattgcactcctgagttctataattcgaatgtcgaatggaactggaactggaaggcaagcttgcctcctctcta

141 D C T P E F Y N S M S N A M E Y G T G G K V T L V A P L

1166 ctctactcaaccaaggcgcaagctggttaaatggggaattgatttagctgttcccttatttcttgactcgttcagcttgatgaaggt

169 L T L T K A Q C V T K W G I D L D V P Y F L T R S C Y E S

1250 gacgctgaaagttgtggaacttgccaacttgatcgaccgcgaaggcattcgaaagaaatggaatgactgacctattctat

197 D A E S C G T C A T C I D R K K A F E E N G M T D P I H

1334 cataaggagaattcg 1348

225 Y K E N *

dp1ORF030

20088 atgaataacgaaaaaattattgaaaaaattaaaaatcttattcaattagcaaatgacaaccgcgagtgacgaagaggggcgaact

1 M N N E K I I E K I K N L I Q L A N D N P S D E E G Q T

20004 gcccttcttattggctcaaaagcttgatgcttaagaataatctcgacttgctcaagttgacaatttgatgaacctaaacgttc

29 A L L M A Q K L M L K N N I A L A Q V E Q F D E P K Q F

19920 gagactctcgaagctgctgggaagaagcaggtcggaataattttggtgggaacgctgaacttggtcatattctcgcgactaatctt

57 E T S Q A V G K E A G R I F W W E R E L G H I L A T N P

19836 aggtgcttttgtattaatcagcgtgatatgcgcttgaaataaaagtcgaataatttttcttcggcgaaaaacaagacgctgaatta

85 R C F C I N Q R D M R L N K S R I I F F G E K Q D A E L

19752 gtgtcttaaaatgatgaggctgctttgtcttattcttcgctacgcttatgaccgacttctcactcgcgaacttctcacaagaat

113 V S K I Y E A A L L Y L R Y R I D R L P T R E P S Y K N

19668 tcataactcaaaaggttttttgcagccttagccattcgattttaaagcaggtggaagaatattcacttattgvtcttacctagc

141 S Y L K G P L S A L A I R F K K V E E Y S L M V L P S

19584 gagcaaaacaaaaaatcgcttcaggagacacattcgaaatttaagaagggaagggaattgacagacactcaactgacttcaactctt

169 E Q T K N A L Q D T F R N L K K E G I D R P Q H D F N L

19500 gaagcgtatattgaaggcggtttctatggcgagaaatcgaaagattatgcgccgataaattttggaaggcggtgaactaa 19423

197 E A Y I B G R F H G E N A K I M P D E I L E G G N *

dp1ORF031

26943 atggcttatcaattagaagacttgctaaaaggctctagatgaaccaactatcaaacagggtgaaggaaattatttgcgaaaacttcg

1 M A Y Q L R D L L K G L D E P T I K Q V K E I I S K T S

27027 aaagaactcgatgctaaaattttcattgcagggcgaggtcaacattttgtacctcaacgacgctttcgtagaagttgttcaacag

29 K E L D A K I F I D G D G Q H P V P H A R F D E V V Q Q

27111 cgcgatgcagcttaacggctcaattaattcttataagaacaagtcgcgacgcttttctaacaagggtcaagataaacggtgatgcy

57 R D A A N G S I N S Y K E Q V A T L S K Q V K D N G D A

27195 cagacacactatccaaacacctcaagagcaactcgacaaacagctctcaacttcgaaaggcgctgctgattactcagctcttcat

85 Q T T I Q N L Q E Q L D K Q S Q L A K G A V I T S A L H

27279 ccgttgattagtcactccattgctccagcagcagacattcttgatgtaaccttgacaacattacggctcgaaagtcagcggt

113 P L I S D S I T A P A A D I L G F M N L D N I T V E S D G

27363 aaagttaaagggtcttgatgaagagtgaaagctgcttcgtgagctctcgtaaatcacttattcaaaagagtcgaaagttccccgcgaa

141 K V K G L D E E L K A V R E S R K Y L F K E V E V P A E

27447 caagaggctcaagctcaagtcgcccgcgggactggaaatttaggaaatccaggctcggtcggtgggtggttcccgaaacctcgt

169 Q R A Q A K S P A G T G N L G N P G R V G G G V P E P R

27531 gaaatcggctcttttgtaagcaacttgctgctgctcaacaaacggcaggagcacaagaacaatcatcattctttaataaa

27611

197 E I G S F G K Q L A A A Q Q T A G A Q E Q S S F F K *

dp1ORF032

52033 atgaaagaagcgaatagactagtttctagctatgtaggattcgaaatgctggactgacgaagaatgtatcaggaaactttgaacta

1 M K E A N R L V S S Y V G F E C W T D E E C I R N F E L

52117 gacctgactcgattcaattcgcttctcttattctgcttattttggatgctcttattccctatgcgaaaaagggtttaaatgcttattc

29 D P D M S I A Y H R Y V G M L Y S Y A K R F G K C L S

52201 cgacatgcacatgaaagcattgcattcagagactatttcaaaatggtttggcaacggttcaaatcaaaccaaggggcccaagtttca

57 R H D I E S I A F E T I S K C L A T F K S N Q G A K P S

52285 acttaccttcacaagactcttcaagaatagaatagctcttagaattatggtaccttaaatgcaccttccatgaatcgaaatgggtat

85 T Y L T R L F K N R I V L E Y R Y L N A P S M N R N W Y

52369 gtgaagtagcgttcgatagcgttctgcacaaatgaagaaggcgacgatttttagtattcctatcgacagttggctattgtgaagac

113 V E V T F D S V S T N E E G D D F S I L S T V G Y C R D

52453 tacggaaaaattgaaattgaagcaagctcttgacttctatgacgcttcttaatacagagatgcttatactctgctcgtcattcaa

141 Y G K I E I B A S L D F M T L S N T B Y A Y I S S V I Q

52537 nacggctcttcagtaagcagcgcagaaattcgccgctgaaattggagtaagcagggctcgctatttagtcagctcgaaaggtcacta

169 N G P S V S D A E I A R E I G V S R S A I S Q S K _ M _ S L

52621 aaaaataaattaaaagattctatataa 52647

197 K N K L K D F I *

dp1ORF033

7670 atggcaagacctaaagtacctcaaatgtatattcgagaagaagaatacagagatgctcaagacgtgcagacactcgatggtgcg

1 M A R P K L P Q I D I R E E I R D A Q Q D V A D S Y G A

7754 attatcaataaagtagtcagcgaattttggaagcagcttcgggttccacttgaccaggcaatggaagaattccaattagttgtg

376

29 I I N K V V D E I V E A A C G S L D Q A M E E I Q I V V
7838 agccaaaatcctgtcattatggaagaccttaactactacattggctatcttccactcttcttatttcgcccagatagggcg
57 S Q N P V I M E D L N Y Y I G Y L P T L L Y F A A D R A
7922 gaaatgggtggaatacaaatggattcaagttctgctatcaggaaagaaaaatcagataatctatacattttagccgcccggaaa
85 E M V G I Q M D S S S A I R K E K Y D N L Y I L A A G K
8006 actattcctgacaagcaagcagaactcgaactgtcatgaatgaagaagtcacgaaaatgcttacaagcgagctacaag
113 T I P D K Q A E T R K L V M N E E V I E N A Y K R A Y K
8090 aaagtccaattaaagctagaacaggccgataaggtattagcatctttaaaacgaattcaaacctggcaactagcagaggttagaa
141 K V Q L K L E Q A D K V L A S L K R I Q T W Q L A E L E
8174 actcagcacaataattcaaaaggagttattataaatgcaaaagacgtagacgtgaaaatgattga 8239
169 T Q S N N S K G V L L N A K R R R R E N D *
dp1ORF034
131 atgagtcacaaactacacgcactgacgtgaattgacaggcgttactcttttaggaaaccaagacaccaaatacagattatgac
1 M S Q N T T R T D A E L T G V T L L G N Q D T K Y D Y D
215 tataatccagacgtccttgaactttccctaacaacatcctgaaaaataattacctagtaaacatttgacggatagtaattcact
29 Y N P D V L E T F P N K H P E N N Y L V T F D G Y E F R
299 tccctttgcccataaacaggacagcctgacttcgagaaatgtttcatttagttacattccaaacgaaaagatgggtgaatctaaa
57 S L C P K T G Q P D F A N V F I S Y I P N E K M V E S K
383 tcaatgaaattgacttattcagtttccgttaaccacggtgacttccacgaagattgcatgaacattatttgaatgacttgcatt
85 S L K L Y L F S F R N H G D P H E D C M N I I L N D L Y
467 gaattgatggaacctaaagtagcattgaagtcattggcctattcactcctcgtgggaatttcaatttaccattcgtcaacaaa
113 E L M E P K Y I E V M G L F T P R G G I S I Y P F V N K
551 gtgaatcctcaatttgcactcctgaacttgaacagcttcaacttcaacgcaaatgaacttcttggaaatgttcaaggtcct
141 V N P Q F A T P E L E Q L Q L Q R K L N F L G N V Q G L
635 ggacgagctattcgatag 652
169 G R A I R *
dp1ORF035
17425 atgcacctaataaggattcgaagatgttgaggacatggaagtccttagcattcgagttcgaacgaagggtgaggacgacgagt
1 M H L M K D S K M L R T W K S L A F E F E T K V R T T S
17341 ggggtgaagttatcgctgctatgaaaaagcagtgacgaggaagatttgaagggttataaaatgaagggtatttatcaacaat
29 G L K L S P A M K T M T R T K I W K G Y K M K V P I N N
17257 catactgaagctgatattgactacaagatattctaaattttgtagcttaccgaaactctcctaaccctcaaatcaaatcact
57 H T E A D I D Y K D I L N F V A Y R N S P N P Q I Q I T
17173 agctggaacgctttgctttcctgctatacacggaatgagctttctataaaggagtttcaataacggactttttgaaagcatt
85 S W N A L L S C Y T R N E L S Y K G V S I T D F P E A I
17089 caaactattgcaagttccttcaactcacctagactcgaaaaacaattgatacacaaaatgaaaacgactcgaaaggattgaggaa
113 Q T I A S S F T H L D S K T I D T Q N E K R L E R I E E
17005 cttcagtcagaataggtcattgttaactgtactatcgacgaacttaaaaaggagttccacgaaatgccggatattgaatcagct
141 L Q S R I G H C N C T I D E L K K G V H E M P D I E S A
16921 atttcttaccagtcaggacagattcttgcttatgaagatgaacttaatttctgctaaactaa 16859
169 I S Y Q Y G Q I L A Y E D E L N F L L N *
dp1ORF036
48808 gtggttagtcgaacgaaaagccgacaaggaatgttgggaatggctagaagctgttcgagcaaatatagtcgaagaagttcgaaac
1 V L V E R K A D K E C W E W L E A V R A N I V E S V R N
48892 ggtccttagcattgttattgcttcgaatactgtcgggaatgggaaaactagctggcggttcgacttttgcaacgctatttagca
29 G L S I V I A S N T V G N G K T S W A V R L L Q R Y L A
48976 gaaactgcacttgacggaagaattgttgagaaggaatgtttagtgctcagctcaactattgactgagttcggcgactataat
57 E T A L D G R I V E K G M F V V S A Q L L T E F G D Y N
49060 tattttcaaacatgcaagaatttctcgaacgttttcgagcgcttaagacttgtgagctattagtcacagcaaatagggtgga
85 Y F Q T M Q E F L E R F E R L K T C E L L V I D E I G G
49144 ggttccttaaccaaggctccttatccttatctgtagacttgggttaattataggggttgacaataacttgcgactatttatagc
113 G S L T K A S Y P Y L Y D L V N Y R V D N N L S T I Y T
49228 actaattatactgacgatgaaatttattgaccttttaggccaaaggctttatagtcgtatatatgataacttcagttggttcagat
141 T N Y T D D E I I D L L G Q R L Y S R I Y D T S V V L D
49312 tttcaggcaagcaatgtaaggaggttggaggtaagcgaatgaatcatag 49362
169 F Q A S N V R G L E V S E I E S *
dp1ORF037
55855 atggtgaagaattgaatctaaaattcattcagttgcataataattctagtagttattgcaacctgtgacaatttatttc
1 M V K K L K S K I Y S V A Y I I L V V I A N L V T I Y F
55939 gaacctttaattgtgaagggaatttttaattcctccaagcagttggtttatgggattcactttcctgcttataaatctataagc
29 E P L N V X G I L I P P S S W F M G F T F L L I N L I S
56023 aagtcaggaagccaaaatttgcaggttctttgatattgggttaggttattccttacctcgttgatttgccttatgcaaaacctta
57 K Y E K P K F A G S L I W V G L F L T S L I C F M Q N L
56107 ccacaatcgcttgcgtgcttcaggagttgcattttggataagtcaaaaagcaagtgcttttatctcgacaagctctcgaat
85 P Q S L V V A S G V A F W I S Q K A S V F I P D K L S N
56191 aaattagactcgaagattgcaaatgctttgtcttagcaacatcggttctattatagacgcaaccatatggatttcttaggactg
113 K L D S K I A N A L S N I G S I I D A T I W I G L
56275 agtcccttgggaattggaacggttcgatatatagatatccgtcagccgactaggccaagttcagttcagtttattctgcag
141 S P L G I G T V A Y I D I P S A V L G Q V L V Q F I L Q
56359 tcaattgcttcgagatatttgaagaagtag 56388
169 S I A S R Y L K K *
dp1ORF038
1350 atgagagttttcraaaccttaacattcgacgcagctcatcaactagttggacattttggaaaatcgcaaatgttcacgggcat
1 M R V S K T L T F D A A H Q L V G H F G K C A N L H G H
1434 acttacaagtcgaattttcattagcagggcgaatttadgaccaggttcgagtcgaagggtggttgcactttatcagctc
29 T Y K V E I S L A G G T Y D H G S S Q G M V V D P Y H V

377

1518 aagaaaatcgaggatcattcattgacagacttgaccagctgttcttcttcaagggaatgaaccaatcgctttagcaaatgca
57 K K I A G T F I D R L D H A V L L Q G N E P I A L A N A
1602 gttgacaccaagcgagttcttattggatttagaactacggctgagaatattgcaagattccttacctggactctcagggagctt
85 V D T K R V L F G F R T T A E N M S R F L T W T L T E L
1686 atgtggaagcatgctcgtatcgactctatcaactatgggaaactcctacaggttgcgcagaatgtacttactacgagatttcc
113 M W K H A R I D S I K L W E T P T G C A E C T Y Y E I F
1770 acagaagacgagattgaaatgttcaagaacgtaacctttatcgacaagacgaaaagattactctccgcgcaaattttagagcag
141 T E D E I E M F K N V T F I D K D E K I T V R E I L E Q
1854 gagcaggataatgggttaa 1871
169 E Q D N G *
dp1ORF039
3306 atgaataaaaagtgcacaccttttggctgttcgaacagctcttatttgcggctctatatgtgacattgacggttcgattttctgct
1 M N K S A T F W L V R T A L I A A L Y V T L T V A F S A
3390 attagttatggacctattcaatttagagtcagtgaaaccttgattcttctacctttatggaaccatagatggactccggggatt
29 I S Y G P I Q F R V S E A L I L S P L W N H R W T P G I
3474 gtattaggaacaattattgcaactcttttccactcttgagctgattgacgttttattcggttcacttgcactcttcttggga
57 V L G T I I A N F F S P L G L I D V L F G S L A T F L G
3558 gtatggcgaatgggtgaaagtgtgtaagatggcaagctcttatttacttacttctgctccaggtcttgcataagcttacttatt
85 V V A M V K V A K M A S P L Y S L I C P V L A N A Y L I

3642 gcgctggaacttgaatagtttactctttaccttttgggaatctgtcattctatgttaggaattagtgaaagcgtattatcgcttta
113 A L E L R I V Y S L P F W E S V I Y V G I S E A I I V L
3726 atttactacttcttatttccacgctggcgagaacaaactcattttagaacactgtaggagcgaaaatgggatttaa 3803
141 I S Y F L I S T L A K N N H F R T L I G A K N G I *
dp1ORF040
7192 gtgagctatctggaataatgttgcaggaagacttttgcgaaggtgcaaaagacttgcgaagatgctttcacggtccgtcta
1 V S Y T G K M F E E D F F E G A K D F E K D A P T V R L
7276 tatgataccactaatggatttgcaggagttgcaaatccctgcgatttatagccgcaactaacttgggacacttgtttattgaa
29 Y D T T N G P R G V A N P C D Y I A A T N F G T I G
7360 ctgaaaactactaaagaagcttcttgcagcttataacatcactgataatcaatgggtccagctatcacgcgcagatggatgc
57 L K T T K E A S L S F N N I T D N Q W P Q L S R A D G C
7444 aaatttattctgcgcgaattttagctgatttccaaaagcatgaaaagatttatgttcccaatttcaagccttgaaaaat
85 K F I L A G I L V Y F Q K H E K I I W Y P I S S L E K I
7528 aaacggctcggagttaaaagcgtcaacccaaacttcatcgatgcagggtatgaagtttcttacaagaagcgtcgaaactagattg
113 K R S G V K S V N P N F I D A G Y E V S Y K K R R T R L
7612 accatttcttccaaaatgttctagatgcagttgagcttctattacaaggagaaaagcaatggcgaagacctaa 7683
141 T I P P Q N V L D A V E L H Y K E K S N G K T *
dp1ORF041
8208 atgcaaaagacgtagacgtgaaaatgattgacccataaacttgaccgattaaaatacacaggtgattgggttgatgtacgaatt
1 M K Q D V D V K M I D P K L D R L K Y T G D W V D V R I
8292 agttctatcactaaaattgacgcccagacgcccagatgtctcaagatgtcgaaaagtgttcaaaaaggctcaagtatattcagtg
29 S S I T K I D A D S A D V S R C R K V L Q K A Q V Y S V
8376 gcggcaggtgaatgcattaaaatttgcacacggttcttgccttgaacttccaaaggataggaagcaatcttgcactccttcc
57 A A G E C I K I A H G F A L E L P K G Y E A I L H P R S
8460 agtctttttaaagaaaactgggtctaatcttctgttcttagcggagtgattgacgaaggttcaaaaagggtgacactgattgggttc
85 S L F K K T G L I F V S S G V I D E G Y K G D T D E W F
8544 tcagtttggatgctactcgtgacgcagatattcttctacgaccaaagaatttgcctcaatttagaatttcaggaaaagcaacttgc
113 S V W Y A T R D A D I F Y D Q R I A Q F R I Q E K Q P A
8628 atcaagttcaatttctgtagaattctttaggaaatfcgggtcgtggaggccatggaagtacaggtgatttctaa 8699
141 I K F N F V E S L G N A A R G G H G S T G D F *
dp1ORF042
48082 gtggcaaggcaagaataggcaattcaggaaaagcctaaaaatgaaattgaactaacattcaagacaagcctaaaactcgttct
1 V A R Q R I G N S G K P K N E I E L T F K D K P K T R S
48166 accttattcaagaaggacgtggcaacaggtcttctcaaaagtgcagcatgatttttcaaatagttgaagcacttaacggaaaa
29 T L F K K D V A T G L S K V E H D Y F Q I V E A L N G K
48250 caattcgaacctaatatgaagcaggtgtcatcttcttattagttcagatgaatttatttcaatttaagtctcatgattat
57 Q F E P N M K Q V S S P F I V Q Y E F I P N I K C I D Y
48334 aactggttcaacttttgcagcactatgaaaaatgttcgaacttatttaaacattgagtcgaacattgaacttctgcgattttta
85 N W F N F S S T M K N V R T Y L N I E S N I E L C R P L
48418 gctgaaagtgtttgtaaatatgaaaatgttcgaaaagattgaacctaaagcgaaggttcataacgggtctcgaactttcaaaaga
113 A E S F V K Y E N V R K R L N L S E R F I T V S T F K R
48502 gcctggattttggacgaactcgaaggaaaacgggttcaaaattcgaaggattttattag 48561
141 A W I L D E L E G K T G S K A P E G F Y *
dp1ORF043
31699 atgactaatattatcacagctgagcaggtttaaagcaacttgcatttcaaatcatcgacttccaggattttcaaaaggtagtgaa
1 M T N I I T A E Q F K Q L A F Q I I A L P G F S K G S E
31783 cctatcctatgttaaaattcgagcagcaggtgtcatgaacctaatcgctaacgggaaaatccctaatcgttctttaggtaaagt
29 P I H V K I R A A G V M N L I A N G K I P N T L L G K V
31867 acagaactgtttggagaacttcgacagctactaaagacaatgctagtctagcatcaattactgaccaacagagaagaagcgc
57 T E L P G E T S T V T K D N A S L A S I T D Q Q K K E A
31951 ctgcagcattgaacaaaacgacacccggtattcaagacatggctgaacttctcgagttatcgagaagcttcaatggtagag
85 L D R L N K T D T G I Q D M A E L L R V F A E A S M V E
32035 cctacttacgctgaagtgcgagtagtatgacagatgagcaacttatgacaatcttcagtgcaatgtacgggtgaagtgaactcaa
113 P T Y A E V G E Y M T D E Q L M T I P S A M Y G E V T Q
32119 gctgaaaaccttctgacagacgaaggaaatgtctaa 32154
141 A E T P R T D E G N V *

378

dp1ORF044
25666 atggtaagtgtttttagcagcagctcctttttgaagttcctgcttcttttagctcgacaagtatttctaaatcgaataag
1 M V S V L I S S S S F L K F L L H F S S T S I S K S N K
25582 gttttcaatttccttgttttctacataagtggtgaaccgataatggcacttaggacattcgaagaattctcactctacgcctt
29 V F N F L V S Y I S G E P I M A L R T F E E S P L Y A L
25498 ttcgatatgtttcgaataactctgttttagatgtaaggctgaacttatgctcacaatggctcacaattaccttgaacgtctgggt
57 F D M F R N N L F R C K V E L M L T M V T I N L E R L G
25414 cgactccttctcgttgggtgttgcagttgttcttttcttctgctcatcaacttcgtcttcttctcactcggttctcttgcaggt
85 R L L L R L V V Q F V L F L C H Q L R L L H S F H L E A
25330 cctcttcttctgttataactcgttctgtaatacaggcaatgctccagctgagatttctgcaagctgagcaagttcttccaaatgc
113 P L V R L I R L L I Q A M L Q L R F R Q A E Q V L P K C
25246 gttccattccttctcgccttttcttcttactga 25211
141 V P I P C P P F P S Y *

dp1ORF045
25340 atgaacgagtgagaagaagcgaagttgatgacaaagaaaagaacaaactgaacaaccaaccgaagaaggagtcgaccagacg
1 M K R V K K T K L M T K K K N K L N N Q P K K E S T Q T
25424 ttcaaggtaatttcgaccatttgtagcataagttcgaccttaccttaaacagattatttcgaacatatcgaaaagggcgta
29 F K V N C D H C E H K F D L T S K Q I I S K H I E K G V
25508 gaggtagagatttctgaatgtcctaagtgccattatcggttcaccacttatgtaggaacaaggaaattgaaaaccttattcga
57 E W R F F E C P K C H Y R F T T Y V G N K E I E N L I R
25592 tttagaataacttctcgagctaaaatgaagcaggaaactcgaagagctgctgctaatcaaacactttaccattcatatcga
85 F R N T C R A K M K Q E L Q K G A A A N Q N T Y H S Y R
25676 attcaggatgagcaagctgggcataaaatctcagggtctatggcgaagctaaagaaggagataaacattgaaaaacgagaaaaa
113 I Q D E Q A G H K I S G L M A K L K K E I N I S K R E K
25760 gaatgggtatcttatatag 25777
141 E W V S I *

dp1ORF046
42774 atgccaatgtggctaaacgacacagcagtccttgacgagcattattacagcgtgcagcggagtgcttactgtcctactaaatag
1 M P M W L N D T A V L T T I I T A C S G V L T V L L N K
42858 ttattcgaatggaatcgaataaagccaagagcgttttagaggatattcttacaactcttagcactcttaaacagcaggtcgac
29 L F E W K S N X A K S V L E D I S T T L S T L K Q Q V D
42942 gggattgaccaaagcagtagcaatcaatcaccaaaatgacgtcattcaagacggaactagaaaaattcaacgttaccgtctt
57 G I D Q T T V A I N H Q N D V I Q D G T R K I Q R Y R L
43026 tatcagacttaaaagggagtgataacaggcctatcaactctcgaccatttttagagagctctctatttttctcgaaggtat
85 Y H D L K R E V I T G Y T T T L D H P R E L S I I F E S Y
43110 aagaaccttggcggaaatggtgaagttgaagccttgatgaaaaatcacagaattacaaataggagggaagatttagatgaa
113 K N L G G N G E V E A L Y E K Y K K L P I R E E D L D E
43194 actatctaa 43202
141 T I *

dp1ORF047
47542 atgaattttgaagatgaaaaacagttcatcgctgcaattgaagaagccggtgaattaaatgctaccaaaaggcgacatggagaaa
1 M K F E D E K Q F I A A I E E A G E L N A T K G D M E K
47626 caagtcaaaagctcttcgtgatgctctaaaagagtagacattgaatcgtctcaaggttaagcattttctgct
29 Q V K S L R D A L K E Y M K E N D I E S A Q G K H F S A
47710 accttccacagcagagcgtctcaactatggcgaagacgcttgaagaattatcgaaaatttagttacgaagccgagagc
57 T P Y T T E R S T M D E E R L K E I I E K L V D E A E T
47794 gaagaaatgctggaanaactttcagggttatcgaaatacaagcctgtcatcaatcgaacttctcgaggatattgattatcac
85 E S M C E K L S G L I E Y K P V I N T K L L E D M I Y H
47878 ggcgagattgaccaagaagcaattcttccagcagttgtcattctgttacagaaggcattcgttttggaaaggctaaaatttag
47961 G E I D Q E A I L P A V V I S V T E G I R F G K A K I *

dp1ORF048
16709 atggaaacaacactttatttctcggttatcttacagcagattggaagacggtcacagaactacactttccactatgaagcatt
1 M E T T L Y F G Y L T A D W K D G H K N Y T F H Y E S I
16625 cctgtaaaagaaactgagaaacaataaggtcactggaatcaatcctaacttgacttagacctaggtcagttattgaaag
29 P V K E T E K Q Y K V T G I N P N L Y L D L G S V I R K
16541 agcgaacttgacattgagtagtattcaagcagtgctctgctgctgaaactggagtcacacttactcgagcatggaagttgatgct
57 S E L D I A V P K A C P V A E T G V T L T R D M E V D A
16457 agaattgaaatcatcaagaatttaactacaagaatcgaacgcttaacgaagaattaaagcaagaatgaacaggtaaacaa
85 R I E I I K K L T T R I E R L N E R I K A R N E Q G K Q
16373 gaaagccgcacctagtagtatcgctagaagattgcgctcgtcaaatgctggaatttatcaataa 16308
113 E S R H L V S A L E D C A R Q I A G I Y Q *

dp1ORF049
44018 atgtttcaaccattttctcagcagcagtgtagccttggctgctcaaagtagaaccaagacttggtttcttctgatatactcgaactc
1 M F Q P F L S E H V A L V V K V E P R L V F F D I L E L
43934 atcttttgataagttccgtttgctcgagcgtaccagaaccagtagcattcttctgcccagccaagttcttctcagcgggtg
29 I F W I S S V C S V P E T S S I F L P A K F L L S R L
43850 agcatttgcgttagtcgaagcagtagcgttagtagtgacgtgcatagtcaccaacgctcatcgtggtcgttgacggaat
57 S I C V S Q A I D V V V R L T C I V P T L I V V D G N
43766 tccgtcgttaggctagttgcagtgatgatgttatcactgtcaatgaacatccctgtatgacctccagcgcctgctgtagcacc
85 S V V G V V A V N D V I T V N E H P C M T S S A - C A S T
43682 tttgcgtcccgagatgaagattgcgctcgttagcatcccgagcagcattttcactaattag 43620
113 P A S P D E D V A S F S I P R S I P T N *

dp1ORF050
15081 atgaacaatcagcgaagcaaatgaacaaacgaatcgtcgaacttcggaagactatcaacgtgcaagggtcgaataaaactc
1 M N N Q R K Q M N K R I V E L R E D Y Q R A R G R I N F

15165 cttcttctgttaaggaccacggcgaagaactcgaaaacccctgaagcccttctgtgggatacattgacaatcttagtgaatgtttt
29 L L A V K D H G E E L E N L E A F V G Y I D N L V E C F
15249 cctgaaagccaacgaatgtcttgaggctattgtattagatgacccctccagtcactaatgcggccgctgaaattggataccac
57 P E S Q R N V L R L C V L D D L P V T N A A A E I G Y H
15333 tatacatgggttcaccaacttcgagacaaagcagttgaaacacttgaagaaatttttagatggggataacattattcgctctaaa
85 Y T W V H Q L R D K A V E T L E E I L D G D N I I R S K
15417 cacggaatcgaattaaaggagaaacttgatgaattatattggttaaagtcattctagtttag 15476
113 H G I E I K E K L D E L Y G K S H S S *
dp1ORF051
29765 atgagttatgacgtgaattatgttaagaatcaagttcgttagagccattgaaaccgctcctactaaaaatcaaggtacttcgaaac
1 M S Y D V N Y V K N Q V R R A I E T A P T K I K V L R N
29849 tcttgggtcagtgatggatgatggaggaaagaaaggataaaagcgaatgaagtcgttagcagacgacccctgtttgttttagttgat
29 S W V S D G Y G G K K K D K A N E V V A D D L V C L V D
29933 aattcaactgttccgtgaccttttagccaattctactgacgagggaataatttttgcctaaatggagtgaaatttttcttcta
57 N S T V P D L L A N S T D A G K I F A Q N G V K I F I L
30017 tatgatgaaggcaaaatcattcaacgagccgatactatcgaaatataaaactcaggaagacgggtacagggtagtagaaccac
85 Y D E G K I I Q R A D T I E I K N S G R R Y R V V E T H
30101 aatcttctcgagcaagacattttgtagaacttaaatggaggtgaacgactaa 30154
113 N L L E Q D I L I E L K L E V N D *
dp1ORF052
30516 atgactaaacgaacgacaatgatggacagattgaagaaattcttctacatttcagctctcgctgctcctatgtcttcagga
1 M T K R T T M M D R L K E I L P T F Q L S P A P M L P G
30600 gctgaatttgacgagcaagatagatagaggcggatgacattgttcttcgatatagtcagatagatgccagcgcaacaaat
29 V E F D E Q D T D R P D D Y I V L R Y S H R M P S A T N
30684 agcctaggaagtttctgttattggaagttcaaatctcagtcctcaactcaattattgttatcgagcaatattagcagaag
57 S L G S F A Y W K V Q I Y V H S N S I I G I D E Y S R K
30768 gttcgaacattatcaaggacatgggtacgaagtaacctatgcagaaactggtagactacttcgacacaaatgctttcttagatac
85 V R N I I K D M G Y E V T Y A E T G D Y F D T M L S R Y
30852 cgactagaatcgaatatagaattccacaaggaggaactaa 30893
113 R L E I E Y R I P Q G G N *
dp1ORF053
50300 atgctaaccattcgaagaatagtatctatcagacaccaaacttgcatcttctactcatttccccgctatatagaagacatcatgc
1 M L T F E R I V S I R A P T C I S L I S P L Y R R T S C
50216 ccgttcttccaagcagttgcaagcattttatcaatgtccacgacttactctgctcaggtcgagccatttagacaaatcc
29 P P F Q A V A S I L S I V H D L P C P G R A I M T I K S
50132 tcaccaggaagtaagcctccaagcagctcgtccaatagttcaaacctgtcgatatccaagtccttcacogtcatggtttcta
57 S P G S K P P S T S S N S S N P V D I P S L S P S W F L
50048 atagtattcgccagtccttagtcgaagtttagcatttcgagcaatgtcttagtcgagcctacgaatttagagcgaattgaaagttct
85 I V P A Q S S R S L A F R A M S S P P T N L B R L K S S
49964 tctagtttggaaattatattcgcaatcgcaatgttactatctacttga 49917
113 S S F G I I F A I A M L L S T *
dp1ORF054
14423 atgtgtgaaattgtcaaaacgaacattcaataactagaattttcaatgaagatgaagtggtatgtcgagcctcattcact
1 M C E N C Q N E T F N T R I F N R D E S G Y V D A S F T
14507 tacaaggagattcgcgacaccgacagcagctatttagcaatcgagcggtagaagaaagaccgtgacagcccttttagtcgtaca
29 Y K E I R D T A A A I S N R A V E K K D R D S L L V A T
14591 gttatgggtcttcccggttctcagcagagaagatttaggcaagagactttgtattgcaaatctcgattggaagcatttcgtgaa
57 V M A L P V S H A E D L G K R L C I A N S R L E A F R E
14675 gctgttcaagaggtcttcgagaatgaaaggtgaagatttaaggagcgttatcttaggtcttatcgagcgttgacaaaaaatt
85 A V Q S A L E N E K A E D L K D V I L G L I D V D K K I
14759 ggcaaccttgcatctgaatttagttgaatcaggagcattataa 14800
113 G N L A L Q L V E S G A L *
dp1ORF055
27627 atgcctaattgtgcgagtttaagaaaactgattttatcaaaccactcgaagcattgtcgcaattcctgaccactacgttgccttg
1 M P N V R V K K T D F N Q T T R S I V A I P D H Y V A L
27711 gctgctcaaattccagctaccgcagcaactcaagtagggaacaagaatacatcttgcgggaacttgctgaaatgctact
29 A A Q I P A T A A T Q V G N K K Y I L A G T C V K N A T
27795 acatttgaaggacgcaaaactggactcgaagtagtatctaccgggtgaacaattcgacggagttatcttcgctgaccaaagagtg
57 T F E G R K T G L E V V S T G E Q F D G V I F A D Q E V
27879 tttgaaggtgaagaaaagtaacctgacagatttagttcacggattcgtcaaatatgcagcccttcgaaaagttggcgatgct
85 P E G E E K V T V T V L V H G F V K Y A A L R K V G D A
27963 gtgcctgaatctaaaaacgcaatgattctgtcggttaaatag 28004
113 V P E S K N A M I L V V K *
dp1ORF056
19151 atggaataataatggaagttatccattttcaaaactcatgtattaaacaagtagacgatgaaaaagaggctcctgttcgaa
1 M E N K W K V I H F Q N S C I K Q V D D E K R R L L F E
19067 gttccaggaactccttctctacaaagtttgggtgaaatgagcttagttaaaaattgaaacacgcgcaggaaatggctattat
29 V P G T P Y R L Q V W V K M S L V K I E T R A G N G Y Y
18983 aaaaggctagtagtccaagacgattttgtattttatggtaaggagtcattagatgggtacttaattgacgcacccataactggc
57 K R L V C Q D D F V F Y G K E S I D G Y L I D A T T I T G
18899 aaatcttggcggaatattgtgagcctatgaacaggtatattctcgaactattgcatcgaggaagcagctgaactgaacaga
85 K S L A E Y C E P M N R H I L E T I A S R E A A E L N R
18815 gctaaaaagcaagaccaacagaaatggagatactag 18780
113 A K K Q D Q Q K W R Y *
dp1ORF057

380

9859 atgcacaaatctctatttggacctaagctagtgctgctagttcaaggcgcaagaaaagaacgggtccaaaacctaaccctaa
1 M Q K S L F G P K L V P A S S R R K K R T V P K P K P K
9943 atcgatgagcaagtggtgagcttatgaaccgcagagagcgctcaagtgcttgcacatctattattattttaaagac
29 I D E Q V V E L M N R R R E R Q V L V H S C I Y Y Y F N D
10027 tcaattatagcagacgggcagctatgacaaatggagccacgaactatattctcttatagtttcgacccctgatgagtttcgacag
57 S I I A D G Q Y D K W S H E L Y S L I V S H P D E F R Q
10111 actgttctctataacgagtttaaacagtttgacggaaatctggaatgggtcttccatacagactgtcagtttgcgttaagggtc
85 T V L Y N E P K Q F D G N T G M G L P Y D C Q F A V R V
10195 gcagaaaggcctttaagaaaatga 10218
113 A E R L L R K *
dp10RF058
15633 atgacatcacgcgcatacaaaaccaattcccacgcgcagagctagtgctaaacaagagaaggcagttgctaagcagttgggagga
1 M T S R A Y K P I P T R R A S A K Q E K A V A K Q L G G
15717 aaagracagcctaattcaggagccactgactactacaaagggtgacgtcgttaacagactcaatgcttatagaatgcaagcagtt
29 K V Q P N S G A T D Y Y K G D V V T D S M L I E C K T V
15801 atgaagccacaaagtccagtcagcttgaaaaaggaatgggttccctaaaaaatgaacaggaaagggttcgctcaaaaactcgactat
57 M K P Q S S V S L K K E W F L K N E Q E R F A Q K L D Y
15885 tctgctatcgcttctcgacttgggtgacggagcggaacagtatatagcaatgtctataagtcagttcaagcgaatattagaggat
85 S A I A F D F G D G G E Q Y I A M S I S Q F K R I L E D
15969 agaaatgataacccctatttaa 15989
113 R N D N L I *
dp10RF059
30154 atgctctcagcctgaattagttggaagcctgaagaattttgtagtaactgtgaacgggtatcgaaacaagtttcaagtcgctgct
1 M S Q P E L V W K P E E F V S N C E R Y R N K F Q V A V
30238 ataacagctctgcgaagtcgctgctactaagatggaagaatacgcgaagacgcagtcctatttggacagacggtacaggggaatgct
29 I T V C E V A A T K M E E Y A K T H A I W T D R T G N A
30322 cgacagaaactcaagggaagctgcttgggttaagcgacagacaaatcatgatagctgtatcacatcacatggactacgggttt
57 R Q K L K G E A A W V S A D Q I M I A V S H H M D Y G F
30406 tggctagaactagctcctggtcgaaaatacaaaattctcgaaacaggctgtagaagacaatgtcgaagaacttttttagagcgttg
85 W L E L A H G R K Y K I L E Q A V E D N V E E L F R A L
30490 agaaggttattagactag 30507
113 R R L L D *
dp10RF060
38070 gtgatagctgtatctgctatccctactccgctcttttccaggttacaccgctcgactccatcacgcccaggagctcccggtaaacct
1 V I A V S A I P T P L F P G T P S T P S R P G A P G K P
37986 gcgtcaccttttagaccttcttagtcgaatccatgtaagtcgtcaggaactaattcgctcgggttcttatttagtattaaaggaca
29 A S P L G P S S R I H V K S S G T N S L G F L L V L R T
37902 ccaatgtatttccagattctgcattaaaatttagctccctaaaatgtcatctcggtatcttaatacaacttgggactccattaca
57 P M Y F P D S A L K L V P K M S S A Y L I T T W D S P T
37818 gtttccctgaaaggactccttcgctcctcatttagcaagtcctcaagtccttttcgaggggtcttggaaaatgatagtagag
85 V S P E R T P S P S S P S K S I K S F R G S W K M I V E
37734 tttgaaaggctcgtcgtag 37717
113 F E R S S *
dp10RF061
19475 atggcgagaaatgcaagattatgccgatgaaattttggaaggcggttaactaaaatgaaattcgaagtttattctgcgcgacta
1 M A R M Q R L C P M K F W K A V T K M K P E V Y S A R L
19391 tttgacgaagaggcgacatagataggtatcgtgaagcactagagaaagtggaaatgctcgttacttttggaaattgataact
29 F D E E A T Y D R Y R E A L E K V G N V A Y F C E I D T
19307 ggcaaccttgtaactgaactcgagctagacagcctagatgacctaactcgcgcttccaaatgtagtgggaactggacataaatta
57 G N L V I E L E L D S L D D L I A L S N V V G T G L K L
19223 tcacggccttatagagaagataagccttttcaatttaggattgttgacgggtacacggaataa 19161
85 S R P Y R E D K P F Q L W I V D G Y M E *
dp10RF062
45284 gtgagaagcttcaatcaattccattcggtgtcaatatcttcttcttgacgagtttataaattccgctcaatcgccattcgta
1 V R S F N Q F H C G V N I F F L D E F K N S V N R P F V
45200 agatgcaggagcaatagatgcaagaatttcttttgggtcttctgtcaaccttttgcgcgaactccaatagaaacaccttttgcg
29 R C R S N R C K K F L L V F C Q P F C A N S N R N T P S
45116 agtttcttctgtagtaacgaagttcttcttagagcgataggcgatgtaagactgagtgacgattcgagtagacgcaggaaaggc
57 S F P D S N E V L L R A I G D V R L S D D S S R R R K G
45032 ttcaacaattcgactttcaagagccttagtaatcgccatcacgctttcttttttaggagcagggttttcgaacagtagatttctc
85 F N N S T F P K S L S N R H H A F F F R S R F S N S R F L
44948 actaactga 44940
113 T N *
dp10RF063
47200 atgaaattcactgaaggaaaaaattgggtataaagttggagagatatgtcaaatgttgaaacgctctctattctacgattaatgtt
1 M K F T E G K N W Y K V G E I C Q M L N R S L S T I N V
47284 tggatgaagcaaaagacttctgctgaagaaaaatacacttctccggttgggtcttctcctgaacctagaacagaccttgaccat
29 W Y E A K D F A E E N N I H F P P F V L P E P R T D L D H
47368 cgtggttctcgattctgggatgacgaaggcgtgaacaaactcaaacgatttagggacaacctaatgctgggtgacttggcattc
57 G S R F W D D E G V N K L K R F R D N L M R G D L A F
47452 tacactcgaactctttaggggaaaactgaaagggaagcaattcaagaagatgctaaagcatttaaacgtgaacatggattggag
85 Y T R T L V G K T E R E A I Q E D A K A F K R E H G L E
47536 aattaa 47541
113 N *
dp10RF064

381

29108 atggctacattgaaagctcttagcaccttaacgtttccggagcagtagtcattcagggtcgggtattttcttgcctgaagcg
1 M A T L K A L S T L I V S G A V V H S G S V F S C P E A
29192 cttgcttcgtcttaattgaacgcaattttcgcttcgagattaaaggcggctgaagatggagaaacggtagaaactgttccctcaa
29 L A S S L I E R N F A F E I K A A E D G E T V E T V P Q
29276 acaattgaatcagttgaagaaattgacgaagttgaacaaatgcgcgaagagatgcggctaaaacggcttcctgagctcgttgaa
57 T I E S V E E I D E V E Q M R E E Y A A K T V P E L V E
29360 ttagcaagagctaatggaattgacattttctcaattttctgaaaaagcgaatatatcgacgctttaactaagtacgaactagga
85 L A R A N G I D I S S I S R K S E Y I D A L I K Y E L G
29444 gagtaa 29449
113 E *
dp1ORF065
51497 atgcagtttgcataacctacatcaaacatctcgatgagctcgctccgtcaatttccggttcatacatataaggatgaataaacgg
1 M Q F V I T Y I K H L D E L V R Q F P F I H I R M N K P
51413 gtatttcaagttcctcttcaggaatgattttatgctcgactttttcagttctccattttcttcgaaacgcttcagggtgac
29 V F I K F L F R N D F M L D F F S S P I S S K R F R A D
51329 gccttgcttaactacttcgctagatgttccaaaattccttttcagccactggtttccatagaacctccatcgtttcgacctaa
51246
57 A L P N Y F A R C S K I P F Q P L V S I E P S I V S T *
dp1ORF066
28898 gtgaccaactgcgtcagggtggaagcaataccactttacgcgtcgtaactcaagttgaactgacgaatgttaccacgtcaggaag
1 V T N C V R W K Q Y H F T V V N Q V E L T N V T N V R K
28814 cttgtcagcgtcagcgaactgagcaattttcttagagtagacagcgatttgaagacctgttttttcagcgatgaattttctcagc
29 F V S V S E L S N F L R V D S D L K T C P F S D E F L S
28730 gtcacttgcaagaagcaagaagttttcccaagaaccttgaacaccaattgcaagagctttcttgatagagctcactcttagcat
57 V T C K K Q E V F P R T L N T N C K S F L D R V T L S H
28646 ttgggtataagtgttcggttcaagaccattcgagtagggcggaacacctgtacgattttcgatgtcatccattgctgctaa
28566
85 L V I S V S V Q D H S S R A N T C T I F D V I H C C *
dp1ORF067
45061 gtgacgattcgagtagacgcaggaaggcttcaacaattcgactttcaagagccttagtaactcgccatcacgctttcttttta
1 V T I R V D A G K A S T I R L S R A L V I A I T L S F L
44977 ggagcaggttttcgaacagtagattttctcactaactgaaccaactttcttcgggtgttcttcaacttcagggaattttctctca
29 G A G F R T V D F S L T E P T S S G C S L T S G I S S S
44893 aggaacttcttttttaggtttgggaacgactctaccttttcgagcaggtcgagcaactgcaggagcagccttttttagcaggttta
57 R T S F L G L G T T L P F R A G R A T A G A A P L A G L
44809 gcagcttcttcttttttaggtttcagtttctcatttctgtgtaccaacgttcgagagttgaagctgaaaggtga 44735
85 A A S S F L G S V S S S I V Y Q R S R V E A E R *
dp1ORF068
29451 atggcagctcaaacggacattgaacttagtcaaaatcaatcatcgataacgataattctccgtcaccaatgactgaccaaaagatc
1 M A A Q T D I E L V K I N I D N D N S P S P M T D Q S I
29535 tcagctcttttagacaagcataaactctgcgcctatgttagttatatgatttgccttaattgaagaccgggaatgacgtggtaacc
29 S A L L D K H K S V A Y V S Y M I C L M K T R N D V V T
29619 cttggacctatcagctctaaagggtgacgcagactactggaacaaatggcgcaattctattatgaccaatataagcaagaacag
57 L G P I S L K G D A D Y W K Q M A Q P Y Y D Q Y K Q E Q
29703 cttgaaactgatgaaagtcgaacgctggttcgacactcttaattgaaagggctgatgggacatga 29768
85 L E T D E K S N A G S T I L M K R A D G T *
dp1ORF069
20411 atgaaactttatcacgccactgattttgataactcttgtraaaattctagctgaaggattgaagccttcagctggagttatttac
1 M K L Y H A T D F D N L G K I L A E G L K P S A G V I Y
20327 ctgacgaagatttatgaaaggctctagccttttatcgcttcaaatgttgatactattgtcgttctcgaacttgaagtagat
29 L A E S Y E K A L A F L S L R N V D T I V V L E L E V D
20243 attgaaaaatgtactgaaagtttcgaccataatgaaaagatgttttgcagcctattttcattttcgacacttgcgcgcttgagct
57 I E K C T E S F D H N E K M F C S L F H F D T C R A W T
20159 tatgacaagacaattgaagtagacgacattgacttttcgaaagctcgaaaatatgatagaaagtga 20094
85 Y D K T I E V D D I D F S K A R K Y D R K *
dp1ORF070
15973 atgataacctttatataaataaacagtggaaggacagttactccaattaaagggtcagccatgcaactgtacgcagacettatt
1 M I T L F K I N S E G T V T P I K G S A M Q L Y A D L I
16057 cctatacaagaggacgatatacagttcgtttgatataactggacttgacctattgttcgagaaaacgtacttgagctcatttca
29 P I Q E D D I Q F V D I T G L D P I V R E N V L E L I S
16141 cggagccgtgtaggagtttcaaaataggtacaaacctcgaccagaatgatgtcgacgattttctacagcagcgcgaagaagaa
57 R S R V G V S K Y G T N L D Q N D V D D F L Q H A K E E
16225 gcgctcgactttgctaactacctaaccgaagctcaaaagtcacaaagcaaaataaataag 16284
85 A L D F A N Y L T K L Q S Q Q K Q N K *
dp1ORF071
38904 gtgaaacaggtcctagagggttcaagggtcttcaagggtcctcaagggtcctcaagggaattcctggacctgcaggagctgacggac
1 V K Q V L E E P K V F K V L K G F K E F L D L Q E L T D
38988 gttcgcaatatactcacctcgttttctctaatagtcacaaacgggtgagggatttagtcatactgacagcggagcagcatagctcg
29 V R N I L T S L S L I V Q T V R D L V I L T A D E L H T S
39072 gtcagtatcaagatttcaatcccgctccattcaaaagacccctgcagcctatacatggacgaatggaaggggaatgacggagctc
57 V S I K I S I P S I Q K T L Q P I H G R N G R G M T E L
39156 aagggtatcccggaagccaggcgagcggtaagactaattatttccatag 39209
85 K G Y P G S Q A Q T V R L I I S I *
dp1ORF072
51045 atgttctctcgtcttcaagttgtctcgaaagtttttcaattatttgttcaggagtcgcttcaatttgaagaccatttactttca

382

1 M F L R L Q V V S K V F Q L F V Q E S L Q F E D H L L S
50961 tcaaaatgcttcaactccttcccttgaacttacttcgaagacgagcagtcgacctagaggcttttgccttcaatggagagct
29 S K C F N S F P C N L T S K T S S R P R G F C F Q W R A
50877 ttccgctttttcagttccttcttcgcttcttcttgaatcctataagagtattaggtccagtttcaacgctcccatatattc
57 F A F F S S F F A F L F E S Y K S I G S S F N V P H I F
50793 gatgatttttcggtccttcgcatatcgggttttaacgacagatag 50749
85 D D F S V F A I S V F N D R *

dp10RF073
14262 gtgaacgcttgccggaagaatacgcagcaagaacttggaacattatcactgaagcagaatacatcaagcgaacagaaaaacct
1 V N A C R K N T T K K L G N L S L K Q N T S S E Q K N L
14346 aagcagttgcaaaacctactcgaaaaaactccagcgcttctcgtcgccctcgcccttaaaagaaagggtgaaataaaatgtgtg
29 K Q L Q N L L E K L Q R L L V A L A L K R K V E I K C V
14430 aaaattgtcaaaacgaacattcaatactagaattttcaatgaagatgaaagtggtatgtcgacgctcattcactacaagg
57 K I V K T K H S I L E F S M K M K V A M S T P H S L T R
14514 agattcgcgacaccgcagcagctattagcaatcgagcggtag 14555
85 R F A T P Q Q L L A I E R *

dp10RF074
32298 gtgcgcaaaagaaaaactccaggattgcaaatgcttattggagtgactattttcagtcgctcctcttttgcctatagaaaggaaa
1 V T K R K I Q D C K C L W S D Y F Q S L L F L Y I E R K
32382 ttacatggattttgggtcaattgcagcaaaaacttggtatctcacaacttcacaagtcatttaaatcttgcctcaaatgtg
29 L H G P W V N C S K N D F G Y L K L H K S I K S C S K S
32466 agcgcaacggctcgactagagcttccgaagtcctttcaaatgggtctgtcttaacaggattagggaaggacttacgactgc
57 S A T A R T R V F E V L S N W F C F N R I R E R T Y D C
32550 ggttaccccttcttcttgggttgcagcgccctctattaa 32591
85 G Y P S S Y G I C S R L Y *

dp10RF075
22447 atggcaaaagttttgctcgttggaattccgctcatggcccaaaagggaatgaaagagccatcgatactgttttctgaaacgaatg
1 M A K F C P L N S V M A Q R E N E R A I D T V F P E R M
22363 gaaccgtctgctatgcagatcgaaggttcgaaaagggtgagcccttgcaccatgttaggagctggagttgttcttacta
29 E P S A M T I S K V R K G E P F V H H V R S W S C F L L
22279 aaaggacgaagttgaacttaggttagtttattctcaggcttattgtcattatcagtcactcctttaatgttaggaacctgtgc
57 K G T K L N L G S L F L R L I V I I S H S F N V G T C C
22195 gtcactaaattcttgcaaacggcttgagctgcttattctag 22154
85 V T K F L P N G L S C F I *

dp10RF076
5728 gtgagagcatttttctcactcacgtcttcgagcaagtggtcgaatgtagggtactcttctatcttctgtaacaatatcaatattg
1 V R A F S S L T S S S K W S N V G Y S S S S V T I S I L
5644 tactcaccattccaataacttttagcgaagattcttcaggaactaatgtgacggttgcggcggtggttcttcttacta
29 Y S P F P I T F S E D S S G T N V T V A A V V F S T S P
5560 ccaactgctctgtttcacaatcacgtcaattcaacatcgctgtcgataatgcacggaaggtttgagccatcatagcgt
57 P N C S A F T I T S I S T S L S I M H R R K F E P S Y A
5476 gtaaacatgcgcatttcgctgcacaaaaatatgccaatag 5435
85 V N M T H S P S P K I C Q *

dp10RF077
14800 atggaacgaataaagacgctatttcacgtgattatgtcaacggcactcatttagaagtagcagctttgttcgataccggtgat
1 M E R I K T L F H V I Y A N G T H L E V A A L F D T V D
14884 gattatgatgagttatagaggacatccagggttatattgataccctgacctttataatcaaaaggagcatttcttcaagtcct
29 D Y D D V I E D I Q G Y I D T P D L Y N Q R S I R M A P
14968 tacaactctgacatcaatgggtgacgtattgtctactgacattttactacgactagatgatattatcagtcgacgcaactgt
57 Y N P D I N G D A I A T D I L L R L D D I I Y V D A T C
15052 gaaactattaaatcagaggagcctattgcata 15084
85 E T I K Y E E P I A *

dp10RF078
17507 atggcaacagtaaaaggaacagtaaaatttgacggagctcttgtaactatcttcgactacgacgatttagagtggaaggat
1 M A T V K E T V K F D G R L V T I F D Y D D L E W E G Y
17423 gcacctaatgaaggattcgaagatgttgaggacatggaagtccttagcattcgagttcgaacgaagggtgaggacgagtggtg
29 A P N E G F E D V E D M E V L S I R V R N E G E D D E W
17339 gtgaagttatcgctgctatgaaacgatgcagaggacgaagatttggaagggtataa 17280
57 V E V I A C Y E N D D E D E D L E G L *

dp10RF079
35288 atggaactgataccattgataaaactcgaacaagggtgacccctgcgcttaccatttgcagcgaatccagtaaccttagaa
1 M E L I P L I N P R T R L T P A L T I C P A N P V T L E
35204 acaattgaagttcccatgctgccaattttagagacagctgaaccaatcattgacccaataaccactaatgaagtttcgaatcagg
29 T I E V P M L P I L E T A E P I I D P I P L M K F R I R
35120 ttccgacctctgaaacctctgtcccacaagctagcaatcttgctaactaatgatgaaagcatgttccagctgtcgataaa
57 F A P P E T I C P T K L A I L L T N D E S M F P A V D K
35036 agtgagcgagaagtgaaagcaatccctga 35007
85 S E P R S E A I P *

dp10RF080
42490 atgttgaaccttacaatacgcgcaaatgttgccagagttcactattggacaaggagctgaaagaaacttgtdaaaacaacy
1 M L N L T K S R Q I V A E F T I G Q G A E K K L - V K T T
42574 attgtgaacattgatgcaaacgagatcaccgtctctgaaactcttcagaccagacttgatgtcggaacctgcagaa
29 I V N I D A N A V S T V S E T L H D P D L Y A A N R R E
42658 cttcgagctgacgagcaaaacttcgcaaaactcgttacgcaatcgaagatgaaattctagctgaacagtcgaagactgaacaa
57 L R A D E Q K L R E T R Y A I E D E I L A E Q S K T E T
42742 gcttcaacagctgaataa 42759

85 A L T A E *
dp1ORF081
55466 atgttcagggaacagtatcgtccatctgttggctcgtcgaagttaaaggggtcgaaatcttcgttcttgcgtcgcgatara
1 M F R N S I V H L L V C V K V K G V E I F V L A S V D I
55382 ctcgaaactcgtattcaggaagactcatatcaggaagccttcttctcgcacggtagctgttgaacatatcccaagtcctgcgc
29 L E L V F R K T H I R K P S S S T G S C L N I S Q V L R
55298 ctgctgttgaacgaatatgatagcttgcacttttagggaactcggtagaagaatcttcaataaacttattcgtctctttgac
57 L L L N E Y D I V C H F R E L G E E I F N N L I R F F D
55214 agatacattcatctgctcagcgattga 55188
85 R Y I H L L S D *
dp1ORF082
44728 gtgaacttcaccttttcagcttcaactctcgaacgttggtacacaaatggaagatgaactgaacctaaaaaagaagaagctgcta
1 V N F T F Q L Q L S N V G T Q W K M K L N L K K K L L
44812 aacctgctaaaaaggctgctcctgcagttgctcgaactcgtcgaaggttagagctgctcccaaacctaaaaaagaagctcctg
29 N L L K R L L L Q L L D L L E K V E S P P N L K K K S L
44896 aggaagaatctcgaagttgaaggaacagcgggaaggttggttcagttagtgagaaatctactgttcgaaaaactcgtccta
57 R K K F L K L R N S R K K L V Q L V R N L L F E N L L L
44980 aaaaagaagcgctga 44994
85 K K K A *
dp1ORF083
35974 atgccttcagggtttttaaatcctgagtccttaaatcctcgaaggtgagtcctacatatcttagcacgggtgcacctttgtcg
1 M P S G F L N P E S L N P A K V S P T Y S S T V A P L S
35890 acaaggtcaattcgtcgaacaaatagcgtctgtctgtagccatctatttctccttaccggtgttacaatgttaccacacctg
29 T R S I P S T N S V C L L A I Y F S F T V L Q C Y Q T L
35806 acagagttctcttacttctattatacaatctctcgcagcttctgcaacgtcgtcattgttctgaactacgattgttccaatgt
57 I E F L Y F Y Y T I L S T V C Q R R H C F E L R L F Q C
35722 tga 35720
85 *
dp1ORF084
15445 atgaattatatggtaaaagtcattctagttagtgctcttctgactgtcagccttttgcattgacttgcctcaatgggtttatttgggt
1 M N Y M V K V I L V S V F V L S A F C M T C S M V Y L V
15529 acaggttaagcaagaggaccacgttagtaccgtcgccttctgatttggcgctctcgaagctctcggcggttctattcgcacactc
29 T G K Q E D H R S T V A L V F G A L V S S A A P Y S T L
15613 tttatcctcgcctatctgcatga 15636
57 F I L A Y L P *
dp1ORF085
10847 gtgatgactataatcaaggacttttctcagccttctgtagactgtcagcattcctccatttgcgaagtttcccaataaacgaaa
1 V M T I I K D P F E P C D T V T H S S I C K F P N K R K
10763 ggcgtcacgctcataactataaccagctccttcttcttcttctcgtataataaattgaagttgattaacagatgctcgtcatt
29 G V T L I T I T S S F F I F T F D N K L K L I N D V V I
10679 atcaattcgagtaaaagtcacacgggtgaactcgaactgagaaatagtgtaggaatcttttgggggtcagtagacatag 10602
57 I N S S K V K P L N S T E N S V R N L L R V S S T *
dp1ORF086
52760 atatgggaaaagtatcaattcaaaaactcaggaacatttagctcaggggtctaataacagagtttttcacactcgtgcaccacgggtg
1 I W E K Y Q F K N Q E H L A Q G L I T S F S H S L T T V
52844 acagcgcaattgtcactctattgtatgatgacccggaaggcgaagacatggattatttctgtag 52906
29 T A Q L S L Y C M M T R K A K T W I I S *
dp1ORF087
30036 atgatttttgccttcattcatatagaatgaaaattttcactccattttggggcaaaaatttttcccgctcagtagaattggctaaa
1 M I L P S S Y R M K I F T P F W A K I F P A S V E L A K
29952 aggtcaggaacagttgaattatcaactaaacaaaggtcgtctgctcagcattcattcgttcttcttcttcttcttcttcttccca
29 R S G T V E L S T K Q T R S S A T T S F A L S F F P P P
29868 tatccatcactgacccaagagtttgaagtagctttagtttagtaggagcgggttcaatgggtctacgaacttga 29794
57 Y P S L T Q E F R S T L I L V G A V S M A L R T *
dp1ORF088
5040 atgaaaaaagttcaaaacttatcaagaatatctaaaactagttgagttcaaacgtcaactttctttaaattcttcgagaaggaaaa
1 M K K V Q T Y Q B Y L K L V E F K R Q L S L N L R E G K
5124 ataggagtcgatgaagcgggttattcaattattcaccctctatagtttcaacaatatcgaggaacctccttcttcttcttcttcttcaaa
29 I G V D E A V I Q L F T F Y S F N N I E B P P F I V L K
5208 atgcaagaggtgcggtgaacgggacttatgaagcaaaactcaatatgctttaaagattttaaattatttag 5279
57 M Q E A A V N G T Y E A K L N M L K R P K I I *
dp1ORF089
12495 acgtcaatcatgctcgtatcaatagtcaggtatttagacacaaaatgccttttcaactgcgcgtcagtcattttctcaaaactca
1 M S I M S L S I V E Y L D T K C L F N C A S V I F S N S
12411 acacaattatcaggaagccttttagcaacttgccttgcgttcaatttttagtaaccatcaaaacaaaggtcccatattctcaaa
29 T Q L S G K A F S N L L R L S I L V T I K T S V P Y L T
12327 tccgggaagccttttccacctcgaactcatttagacagaaactccttatcatctcgaacagcgaatattcagatga 12256
57 S G S L P H L D S L D R N S L S S R T A N I R *
dp1ORF090
27037 atgctaaaattttcattgacggcgacgggtcaacattttgtacctcagcagcttttcgatgaagttgttcaatagcgcgagtcag
1 L K F S L T A T V N I L Y L T H V S M K L P N S M Q
27121 ctaacggctcaatttaattcttataaagaacaagtcggcagcgttttcaaacaggtcaagataaacgggtgatgcgcagaccacta
29 L T A Q L I L I K N K S R P L N R S K I T V M R R P L
27205 tccaaaaccttcaagagcaactcgaacagcagcttcaacttgcacaaaggcgtgtga 27261
57 S K T F K S N S T S S L N L Q K A L *

dp1ORF091
43189 atgaaactatctaacgaacaatatgacgtagcaaaagacgtggtaaccgtagtcggtccagcagcgattgcactaattacaggt
1 M K L S N E Q Y D V A K N V V T V V V P A A I A L I T G
43273 cttggagcgtgtgatcaatttgacactactgctatcacaggaaccattgcacttcttgcaacttttgaggtactgtcttagga
29 L G A L Y Q F D T T A I T G T I A L L A T F A G T V L G
43357 gtttctagccgaaactaccaaaggaacaagaagcttcaaaacaatgaggtggaataa 43413
57 V S S R N Y Q K E Q E A Q N N E V E *

dp1ORF092
46989 atgaaaactatctccatattaaggaagacactaaaaggaagcggacaggaacggaagaaaactgcactcgaactagctcaa
1 M K T I S I L R K D T K R K P D R N G R K T A L E L A Q
47073 gagattgatgtcaccttagtgagtagcagagctccttcaattcctgaaaggacggcaaccagaattttaaactcgacaaa
29 E I D M S P S E L A E L L Q I P E R T A T R I L K L D K
47157 ctgctcaacaagagcaatgctcaataatagaaaggtatataaatgaaattcactga 47213
57 L L N K E Q C S I I E R Y I N E I H *

dp1ORF093
45756 atgcaacatagcattaaacaatgtttgaaacttgcttctctgtaactgcaatatcaattgctgttagtttccctaaacct
1 M Q H T I K Q C L K L A F L L T A I S I A C L V F P K P
45672 tgctcatcgctaaaaggaacatggatgctcttgctgctattcgaaacattcaacctgggtgcgcaatggagtagtcttgaaac
29 C S S P K R K H G C S C A Y S K H S T W C A N G V V L N
45588 gaaaactgctcattgcttgaagaagctattcggttctgagagtcattgtag 45538
57 E N C S L L E E A I R F R E S M *

dp1ORF094
8281 atgtacgaattagttctatcactaaaattgacgcgcagcgcgatgtctcaagatgtcgaaaagtcttcaaaaggctcaag
1 M Y E L V L S L K L T P T A P M S Q D V E K C F K R L K
8365 tatattcagtgccgagcaggtgaatgcattaaaattgcacacggatttgctcttgaaacttctaaagggatagaagcaatcttgc
29 Y I Q W R Q V N A L K L H T D L L L N F L R D M K Q S C
8449 atcctcgttccagtccttttaagaaaactggcttaa 8484
57 I L V P V F L R K L V *

dp1ORF095
8877 gtgggaaaactacttcagctctcgacattgtcaagaatgcgcaaatggattttgagcaggaatgggaacagaagactgaagaac
1 V G K L L Q L S T L S R M R K W Y L S R N G N R R L K N
8961 tcaaggaagcgtggaaaatgcgcgtgcattccaaagctagcaagactgctgtcaaggaacttgaatgcaactcgatagcttcc
29 S R K S W K M R V H P K L A R L L S R N L K C N S I V F
9045 aagagcctcttaagattgtatatcttgaccttgagaatacattag 9089
57 K S L L R L Y I L T L R I H *

dp1ORF096
46681 gtgattcataaattcttcaatttcgttgaacttatctcggttttctcctgttaccagggttgcatttgactgtcttcgaaagtat
1 V I H K F P N F V E L I C G F S C Y Q V A F D C L R K Y
46597 cttagcaagaggttcaataaccttttcccaattgctaatatcacgcaggacttctccttgctggatacattcctcgacaatttc
29 L S K R F N N L F P I A K Y H A G L S L L D T F L D N F
46513 gatacatctttcgaaacttgcaagacttgacatcttgagtagttaa 46469
57 D T S F E L A R L D I L S S *

dp1ORF097
39100 atggacgggattgaaatcttgatactgacgcagctatgctcgctgctcagtagtactaaatccctcaccgtttggactatt
1 M D G I E I L I L T D V C S S A V S M T K S L T V W T I
39016 agagaagcggaggtgagtagtattgcgaacgtccgctcagctcctgcaggtccaggaattccttgaagcccttgaggacctgaag
29 R E S E V S I L R T S V S C R S R N S L K P L R T L K
38932 acctgaaactcctctaggacgtgtttcacctatcttggaaactga 38888
57 T L N S S R T C F T Y L G N *

dp1ORF098
43627 gtgaaaatgctccgtgggagtgctaaacgagcgacatcttcatctggggacgcaaaggtgctagcgcagcgctggaggtcata
1 V K M L R G M L N E A T S S S G D A K V L A Q A L E V I
43711 cagggatgttcattgacagtgataacatcattcactgcaactacgcctacgcaggaatttccgtcaacgaccacgatgagcggt
29 Q G C S L T V I T S F T A T T P T T E F P S T T T M S V
43795 ggtactatgcaggtcaaccttactactacgtctatcgcttga 43836
57 G T M Q V N L T T T S I A *

dp1ORF099
38298 atgcaagttcgccatctgctactgaagctccagctggtggatggtctacgcaagttccctaccgtcccaggtggtcagtagttat
1 M Q V R H L L L K L Q L V D G L R K F L P S Q V V S I Y
38382 ggactcgaaacaagatggcgctacactgaccaaactatgaaattggatattcagtttcaagaatggcgagcaggggtcctaaag
29 G L E Q D G A T L T K L M K L D I Q F Q E W A S R V L K
38466 gtgacgcaggtcgtgacggttgcaggaaagacggaatag 38507
57 V T Q V V T V L Q E R T E *

dp1ORF100
1597 atgcagttgacaccaagcgaggtctatttgatttagaactacggctgagaatatgtcaagattccttacctggactctcacgg
1 M Q L T P S E F Y L D L E L R L R I C Q D S L P G L S R
1681 agcttatgtggaagcatgctcgatcgactctatcaactatgggaactcctacaggttgccgagaatgtacttactacgaga
29 S L C G S M L V S T L S N Y G K L L Q V A Q N V L T T R
1765 ttttcacagaagacgagattgaaatgttcaagaacgtaa 1803
57 F S Q K T R L K C S R T *

dp1ORF101
19220 gtgataatttttagtccagttcccactacatttgaaagcgcgattaggtcatctaggctgtctagctcgagttcgattacaaggt
1 V I I L V Q F P L H L K A R L G H L G C L A R V R L Q G
19304 tgccagttcaatttcacaaaagtaagcgacatttccaaactttcttagtgcttcacgatacctatcatatgctcgctcttctg
29 C Q Y Q F H K S K R H F Q L S L V L H D T Y H M S P L R

385

19388 caaatagtcgagcagaataaaacttcgaatttcatttttag 19426
57 Q I V A Q N K L R I S F *
dp1ORF102
4034 atgataacgtgggaatgtttgactgtatcgccgaactcgataaaattcctgggtatttagacagcctaagacacgtgaacagc
1 M I T W E C L T V S P N S I K F L V Y L D S L R H V N S
4118 ttttggagcaccacaaatttcttgggataattatctatcatgctcgagcgaatgggtgagaaagacaagctcttaccctattt
29 F W K H H K F L G I I I Y T C A S E W L R K T S S Y L F
4202 tccatagggagaagacttttaaatggctcaactga 4237
57 S I W E K T L N G S T *
dp1ORF103
49352 ttgaatcatagatatagtaacatcacaaactatttttcttggcagattgtctttctttgtattttgctgcgcggtgtcctattgt
1 L N H R Y S N I T T I F L W Q I V F L C I C C A V S Y C
49436 gcaggagtcgataatgagcgagaggtctcaagataagggtgattcacaagtataagcagaaagaaaagtcagcgtctacttgaca
29 A G V H N E R E S Q D K V I Q S Y K Q K E K S A V Y L T
49520 gtcgatagttcaggagcttggctaggaagtgctccgggagccaaggaaagtcctctctacaatgaaaggacagcatgtagga
57 V D S S G A W L G S A P G A K E S P L Y N E K G Q H V G
49604 aaattgaaagaggtgggagagtgga 49627
85 K L K E V G E *
dp1ORF104
21427 atgagaaaaagagtgattttgaagctaaaaaggttgaactggatgtccttaattcctactctcgaatgggttgagtttttcgaa
1 M R K R V I L K L K R L N W Y V L S Y S R M V E F F E
21343 cttttgaacttttcgaatgggttcgacttttcgaaggattgaggttttcgaaccggttgagtttttcgagcattctcgacttttc
29 L L N F S N G S T F R R I E V P E P V E F F E H S R L F
21259 gacccctttctatgtcgcgacttttcgagtggtttga 21224
57 D P F L C S T F R V F *
dp1ORF105
2028 atgatatgcgcacccaccagttcgaatgaaaatagctctttgacctataaccattccttcaccttgaaattgtaggaccgaaaaat
1 M I V A S T S S N E N S L L T Y N H S F T L N C R T E N
1944 ttcctgatagggcattttctcagggtcgcgaacattgattcgaatcttgctctcttcaggctgattgtattgattaaccattat
29 F H D R H F L R V A N I D S N L A S F R L I V L I N H Y
1860 cctgctcctgctctaaaatttcgcggacagtaa 1828
57 P A P A L K F R G Q *
dp1ORF106
10529 atgaacctcgatcaatgatgtaaactttgaactcgctgtccatagacttgatctagaatcttcaataatgtttcgaacattttc
1 M N L V N D V N F E L A V H R L V S R I F N N V S N I F
10445 taccctattattagaagcagcatcaatttcaataggagagccaagtcctttgttcacatccttcgcgaaaattcgagcagtagt
29 Y P I I R S S I N F N R R A K S F V H I L R E N S S S S
10361 ggttttaccagttccagcgcacacagaatag 10329
57 G F T S S A T T E *
dp1ORF107
10750 atgagcgtgacgcctttctgtttattgggaaacttgcaaatggaggaatgcgtgacagtatcacaaggctcgaaaaagtccttg
1 M S V T P F R L L G N L Q M E E C V T V S Q G S K K S L
10834 attatagtcacatcggttgacatggagccgtttctaatgcactag 10878
29 I I V I T L T W K P F L M H *
dp1ORF108
49447 atgcactcctgcacaataggacaccgcgcagcaatacaaaagaaagacaatctgccaaaagaaaatagttgtgatgttactata
1 M H S C T I G H R A A N T K K D N L P K K N S C D V T I
49363 tctatgattcaatttcgcttacctccaatcctcttaccattgcttgctgaaaatctagaaccactgaagtatcatatatacagac
29 S M I Q P R L P P I L L H C L P E N L E P L K Y H I Y D
49279 tataaagcctttggcctaaaagggtcaataa 49250
57 Y K A F G L K G Q *
dp1ORF109
31632 atgtggttgcgaagtcceaaatagttgattctccttcaactttccagcctttgaaagccttacctgttaaggtagggtcaact
1 M W L S K S Q I V D S P S T F Q P L K A L P V K V G S T
31548 ggttttggagaaatcttcttacctgcttcaactcgaactgcgtcgccggttctctgttccaccgttcaaatcgaatgtcacgcga
29 G F G E I F L P A S T R T A S A V P V P P P K S N V T R
31464 cgaagaaccgctggaagttgtgccacatag 31435
57 R R T A G S C A T *
dp1ORF110
16444 atgatttcaattctagcatcaacttccatgtcgcgagtaagtgtgactccagtttcagcgacaggacatgctttgaactactgca
1 M I S I L A S T S M S R V S V T P V S A T G H A L N T A
16528 atgtcaagttcgctctttctataaactgagcctagggtctaagtacaagttaggattgattccagtgaccttatattgtttctca
29 M S S S L P L I T E P R S K Y K L G L I P V T L Y C F S
16612 gtttctttacaggaatgctttcatag 16638
57 V S F T G M L S *
dp1ORF111
28657 gtgactctatcaagaaagctcttgcaattggtgttcaaggttcttgggaaaacttcttgcttcttgcaagtgacgctgagaaat
1 V T L S R K L L Q L V F K V L G K T S C F L Q V T L R N
28741 tcatcgctgaaaaaacagggtcttcaaatcgctgtctactctaagaaaattgctcagttcgctgacgctgacaaacttctgacg
29 S L K K Q V F K S L S T L R K L L S S L T L T N P L T
28825 ttggtaacattcgtcagttcaactga 28851
57 L V T F V S S T *
dp1ORF112
32207 atgcaaaactgatttaggcaaatctgcttcgacgcagcagccgttgccttatattagatatttgaggagacaagactcctagg
1 M Q T D L G K F D A A V A Y I R Y L Q E D K T P R
32291 tatcctgggtgacgaaaagaaaatccaggattgcaaatgcttatggagtgga 32341

386

29 Y P G D E K K N P G L Q M L M E *
dp1ORF113
17715 atgaaaacagtttaagaagcaatcaaaacattcggtgatgaatggtggtacgaaattatcaacgaaaacggccaaatgattcaa
1 M K T V K E A I K Q F G D E W W Y E I I N E N G Q M I Q
17631 gacggaagaatcgaagacatgggcaatacatggaagaacggctcgaccaagtttaagttcatcaactatggtgacatcgaattct
29 D G R I E D M G E Y M E E T V D Q V K F I N Y G D I E S
17547 gaaattatcaaaactatatatcgcataa 17521
57 Q I I K L Y I A *
dp1ORF114
52952 atgctattggcgaagacggggaacagtcctcctgataaattgtccattatgccaaaacggattccctcgtattgaaaaactat
1 M L L A K T G K Q S I L I I V H Y A K T D S L V L K N Y
53036 ttcttcaactttacaaccatgatacgggaaaagtgaacatgggacggagggcgttcttatggttcaaaagattggtacattta
29 F P N P T T M I R E K L K H G T E A V L M F K R L L H L
53120 tcaataaataatggaagccttctga 53143
57 S I N M E A L *
dp1ORF115
5342 atgagcctccttttttgaratatataatatacacgaattatcgcgagtttgtaagccgttttcaataaatttttaaatctttt
1 M S L L F L I Y I I Y T N Y R E F V K P F L N N F K S F
5258 aagcatattgagttttgcttcataagtcctcgttcacggcagcctcttgcaatttgagtacaatgaaaggaggttccctcgatatt
29 K H I E F C F I S P V H G S L L H F E Y N E R R F L D I
5174 gttgaaactatagaagggtgaataa 5151
57 V E T I E G E *
dp1ORF116
20662 atgaaaattttcaaaactttgctaaaagcacttactaatgaatacctaatggtagtgaacaatgaccaagctgaagcttaggcgca
1 M K F S N F A K A L T N E Y L M V V N N D Q A E V L G A
20578 ggaaatatcgaaaacatttctcaacgggttcgaactttgctaattgttagctgaagcgacagtttttaaaactcgaaaactcagc
29 G N I E N I L N G S N F A N V V A E A T V L K L E K L S
20494 gaagaggaagctattgagtag 20474
57 E E S A I E *
dp1ORF117
24680 atgataacaggtcgtcgcaacatttttaaatcgaagtgaatctcgttaagtcactaatagttttgttcaagttatctgctactgtg
1 M I T G C S N I L N R S E S R K S L I V L F K L S A T V
24596 ataaggtcctttgacatcgcttgcctccttatatgtcattagtcgaatggttcatttaagaataactcgacaaggaaatttgcctcaag
29 I R S L T S L V P Y M S L V N G S L R I T R Q G I C F K
24512 ccggttggggcggtattctga 24492
57 P V G A D S *
dp1ORF118
15023 atgatattatctacgtcgacgcaactttgtgaaactattaaatacagaggagcctattgcatgaacaatcagcgaaagcaaatgaa
1 M I L S T S T Q L V K L L N T R S L L H E Q S A K A N E
15107 caaacgaatcgtcgaacttcgcgaagactatcaacgtgcaagaggtcgaataaacttcttctgctgtaaaaggaccacggcga
29 Q T N R R T S R R L S T C K R S N K L P S C C K G P R R
15191 agaactcgaaaaccttga 15208
57 R T R K P *
dp1ORF119
41054 atggaggttcaacatccccgattcagtcagtcctactcttttcgggcatttcttttagtagacacgacttcagcggttcgacagat
1 M E V Q H P R F S T S Y F F G H F P S R H D F S G S T D
41138 tttaacagggnaaaccttctccaaatcatgtcgaaacttcaagtcacttcaacaatgcttcggcgcttacggatccactat
29 F N R E Q L P P N H V E H S S Q L Q Q C F R R L R I H Y
41222 ccaagcatttcacgctga 41239
57 P S I S R *
dp1ORF120
28387 gtgttgaagcgcaagcagaatacatgcgtatgcaattgcttcaatacggtaaatcactgtcaaatcaactaacgagcgaggctc
1 V L K R K Q N T C V C N C F N T V N S L S N Q L T A R L
28471 aatacacttcagactacaacatggtatgctaaagcaaatatgcagtcactaagaaatggactaaccagctgaaagtgaacctta
29 N T L T T T T W M L S N N M Q S L R N G L T Q L K V T L
28555 tcgctgacatttttag 28569
57 S L T F *
dp1ORF121
39222 gtgcagacggatcacgtgagttcagtttggaaagataataatcaacaatatatgggttattactccgattatgagcaagcagata
1 V Q T D H V S S V W K I I I N N I W V I T P I M S K Q I
39306 gcagggatcgaaactaagatcgatgggttgaccgcttgcgaatgttcaagtgagggtcgaaacgagttccttaattctttat
29 A G I E L S I D G L T A L P M F K W E V E T S S L I L Y
39390 ttgaatttgggttaa 39404
57 L N L V *
dp1ORF122
40402 atgttatttctcttactcctacataccgaatcacgttcctgctgattaaacgagttattgttccgttctaaatcgccgacttg
1 M L F S L S Y I P N H V H V W I K R V L P R S K S A D L
40318 Aatggattgggtaaagatcccggttatcgatgtgaatgaaccttggtgaaggtacataaacttcatcctcgtcggaacaataga
29 N G L G K D P V I D V N E P L R K V H N F I P C G G _ E - H R
40234 aattcggtcacttga 40220
57 N S V T *
dp1ORF123
21327 atgggttcgacttttcgaaggattgaggttttcgaacgggtgagtttttcgagcattctcgacttttcgacccctttctatgct

387

1 M V R L F E G L R F S N R L S F S S I L D F S T P F Y A
21243 cgacttttcgagtggttttcgaggttttcgagcaggttcgacttttcgagaaattgagtttttcgacctctaaattaggttcgatt
29 R L F E C F E V F E Q V R L F E K L S F S T S K L G S I
21159 attcgaaaagtttag 21145
57 I R K V *
dp1ORF124
17891 atggtaaaagttaaagatttgcagtaggaatgaaagttgtaaatgcaaaaggtactgaatttaaagtaactgacctgaaggt
1 M V K V K D L Q V G M K V V N A K G T E F K V T D R Q G
17807 cgtaaaatgggttaagccttagaacgtcttagtgatggacgtattcggttctatgataacgaatcactaatggacgaaaaagtgag
29 R K W V S L E R L S D G R I R F Y D N E S L M D E K V E
17723 gtagtaaaatga 17712
57 V V K *
dp1ORF125
49916 atgtcctcagccgcttccgtttaaattggaacaagtgaattatagatgctcctcttttagcttgcgataaggtattcatca
1 M S S A A S V K I G T S E L Y R C S S F S L S I R Y S S
49832 gtttcgccaatttcgaaaaattcgaatccaggaaaatggcgcagaatagtttcgtcgctcgggaactcttccatctatcgaagaag
29 V S P I S K N S N P G K W S R I V S S S G T L P Y L E K
49748 tgttcttga 49740
57 C S *
dp1ORF126
16136 atgagctcaagtagcttttctcgaacaatagggtcaagtcagtttatcaacgaactgtatatcgctcctcttgcgtataggaata
1 M S S S T F S R T I G S S P V I S T N C I S S S C I G I
16052 aggtctgcgtacagttgcattggctgaccttttaattggagtaactgttcttctcactgtttattttaataaggttatcatttct
29 R S A Y S C M A D P L I G V T V P S L F I L N K V I I S
15968 atcctctaa 15960
57 I L *
dp1ORF127
13511 atgctaaatagctttccattccacgtcgtctgttcttgcgccatttttcagtttcacgatactgaccaactttgcaaggtcgt
1 M L N S P P I H R R C S C A I F Q P H D T D Q L C K G R
13427 gaaatagtgctacgattgcaactgtttccattgggttaaatgtcttccagccttgcctaccatggtatccatttcgaaaaagta
29 E I V L R L Q L F P L G K C L P S L C L P W Y P F R K V
13343 gttgattga 13335
57 V D *
dp1ORF128
4852 atgacagcagttcaacaagttaagttctacttagaagaagcggcgctcactttctaaaagatgttgagtacagtgacaactta
1 M T A V Q Q V K F Y L E E A G A H F L K D V E Y S D N L
4936 gagcaagcaattatgaaagatattcttaaatggaatggcgctcatagagatgagcacgatatgaaaaataacttcacacgaagta
29 E Q A I M K D I L K W N G A H R D E H D M K I T S Y B V
5020 ttatag 5025
57 L *
dp1ORF129
25133 atgaacttttcgctaagcaacttgcgtcactgaagttcaactaatgtacgcagccaccaatcttaccattgaagaattcagta
1 M N F L L S N L R S L K F K L M Y A A T N L T L K N S V
25217 agaaggaagcggacaaggaatgggaacgcatatttggaagaacttgctcagctgacgaaatctcagctggagacattgcctg
29 R R K R R T R N G N A F W K N L L S L T K S Q L E H C L
25301 tattag 25306
57 Y *
dp1ORF130
16789 gtgcttgacttttattcttattatcgatataatcataatataataaaacaagcgtcaaggacgcagaagaggtcaattatgg
1 V L D F I P L L S Y N H N I N K T S V K D A E R G Q L W
16705 aaacaacactttatttcggttattcttacagcagattggaagacggtcacagaactacactttccattatgaaagcattcctg
29 K Q H P I S V I L Q Q I G K T V T R T T L S T M K A F L
16621 taa 16619
57 *
dp1ORF131
43846 atgctcaacgggtcgagaagaaacttggtggcagaaagatgctactggtttctggtacgctcgagcaaacggaacttatccaa
1 M L N R L R R N L A G R K M L L V S G T L E Q T E L I Q
43930 aagatgagttcgagtatatcgaaagaaacagtccttgggtctacttgacgaccaaggctacatgctcgtcgagaatggttga
44013
29 K M S S S I S K K T S L G S T L T T K A T C S L R N G *
dp1ORF132
15304 gtgactggaaggtcatctaatcacatagcctcaagacatttcggttgctttcaggaaaacattcgactagattgtcaatgtat
1 V T G R S S N T H S L K T F R W L S G K H S T R L S M Y
15220 cccacaaagggttcaagggttttcgagttcttcgctggttctttacagcaagaaggaagtttatcgacctcttcgacgttga
15137
29 P T K A S R F S S S S P W S F T A R R K F I R P L A R *
dp1ORF133
8061 atgacttcttcattcatgacaagttttcgagtttctgcttgcgttcaggaatagttttcccgccggctaaaatgtatagatta
1 M T S S P M T S F R V S A C L S G I V F P A A K M Y R L
7977 tcgtatttttcttctctgatagcagaacttgaatccattgtattccaccatttccgctctatctgcggcgaataa 7900
29 S Y F S F L I A E L E S I C I P T I S A L S A A K *
dp1ORF134
498 atgacttcaatgtacttaggttccatcaattcatacaagtcattcaaaataatgttcattgcaatcttcgtggaagtcaccgtgg
1 M T S M Y L G S I N S Y K S F K I M F M Q S S W K S P W
414 ttacggaaactgaataagtaaatccaattcaatgatttagattcaaccatcttttcggttggaatgtaa 349

388

29 L R K L N K Y N F N D L D S T I F S F G M *
dp1ORF135
780 atgaagcagaacttgaataatgctgtaactgttgcaatgttctacggagtcgaagttcaccattcttgaaattgactcgaataatct
1 M K Q N L K M L L M L Q C S T E S S S S P F L K L T R K S
864 actcaagctctagctcttcccttattacaaggaaaaggcgaataatttcacatggaaaatcttacgctgaaatcctag 938
29 T Q A L A L P Y Y K E K A K F H M E N L T L K S *
dp1ORF136
55252 gtgaagaaatcttcaataaccttattcgcttctttgacagatacattcatctgctcagcgattgagttagccccgcggcgctac
1 V K K S S I T L F A S L T D T F I C S A I E L A P R P Y
55168 ataagacctaaaagaacggacttgacagaatttcttcgaagtttcccttctgttagtcgttcgctcggaagtag 55094
29 I R P K R T D L T E F L R S F P S L L V V P S G *
dp1ORF137
37146 atgcttcgaacttgtttgttagcaccgtcaggaggacaaactagtccaacccattcacctgcgtctttgataatattctagcgcg
1 M L R T C L L A P S G G Q T S R T H S P A S L I I S S A
37062 acagcgcctacagaagaacgctgtttcaacttccatggcaagccttctgctagttcataccataatgcgtag 36988
29 T A P T E E A T C F N F L G K P S A S S Y H N A *
dp1ORF138
30662 atgactatatacgaagaacaatgtagtcacccgctatctgtatcttgcctcgaatttcaactcctggaagcataggagcagg
1 M T I S K N N V V I R P I C I L L V K F N S W K H R S R
30578 cgagagctgaaatgtaggaagaatttccctcaactctgctcattgttcgttttagtcattgttcactcctag 30504
29 R E L K C R K N F L Q S V H H C R S F S H V H S *
dp1ORF139
12092 atgataactaaatcactcaacttgtttgacccctcctgataaattcgttcacgcagacacgcgcatttgagcccttttagatacc
1 M I L N H S T C L T L L I N S P T Q T R A F E P F L D T
12008 ttctcgaaacactagatgcttccctcactaaaaggtcatgggctcgaagttcttcgaagacatttctacatag 11934
29 F R K H L D A S L T K R S W A S S S S K D I S T *
dp1ORF140
20562 atgttttgcataatttccgtgcgcctaaagacttcagcttggctattgttcactaccattaggtattcattagtaagtccttagca
1 M F S I F P A P K T S A W S L P T T I R Y S L V S A L A
20646 aagtttgaaaatttctatttatttcccttatttggtttttcttatactattattatacaataatgattga 20717
29 A G E N P I L F S L Y L P F F I L L L Y N N D *
dp1ORF141
42922 gtgctaagagttgttagagatatctctaaaacgctcttggctttattcgatttccattcgaataacttatttagtaggacagta
1 V L R V V E I S S K T L L A L F D P H S N N L F S R T V
42838 agcactccgctgcagcgtgtaataatcgctcgtaagactgctgtgctgttagccacattggcatagattga 42767
29 S T P L H A V I I V V K T A V S P S H I G I D *
dp1ORF142
31898 gtgactgtcgaagtttctccaaacagttctgtcacttcttcaaaagcgtatttagggattttcccgtagcgattaggttcag
1 V T V E V S P N S S V T L P K S V L G I P P L A I R F M
31814 acacctgctgctcgaatttcaacatggataggttcactaccttttgaaaatcctggaagtcgcatgatttga 31743
29 T P A A R I L T W I G S L P F E N P G S A M I *
dp1ORF143
7565 atgaagtttgggttgacgcttttaactccagacgctttaaatttttcaaggcttgaaattggataccataatcttttcacgc
1 M K F G L T L L T P D R L I F S R L E I G Y H I I P S C
7481 ttctggaaatacactaaaattccggcgagaataaatttgcattccatctgcgcgtgatagctggaaccattga 7410
29 F W K Y T K I P A R I N L H P S A R D S W N H *
dp1ORF144
36517 gtgcaaatcaagcgactaacttatttagatacattaaacgagggcatttctcaagattcctaatggaattcaacaattacca
1 V Q I K R L T Y L D T L N E A H S S R F L M E I Q Q L P
36601 ttgaataccgagcagatgacgcagcagcttggaccttacttctcccgctcaagttgaaactgtttctaa 36669
29 L N T E P M T Q Q L G P L L F P L K L N C F *
dp1ORF145
42067 atggaacacagctggagacctaacaagtggaagaggttctatttaagcaagacttcgaacagaataattggcagaaactgttcc
1 M E T A G D L T S G K R F Y L S K T S N R I I G R N L F
42151 ttcaagtggtgggaaccatcactcaacctatggcgacgcatttctattcgaaaactcttgacggcatag 42219
29 F K V G G T I T Q P M A T H S I R K L L T A *
dp1ORF146
51484 atgacaaactgcatgattgcatcacctttccagtcaggaacctcaagggcgaacagttattcttcaaccgtcgaagttgttcgtt
1 M T N C M I A S P F Q Y G T S R A K Q Y S S T V E V F V
51568 ctaagtttccacagtcaggtgaagatgacctaaaacggaatttctttatggccaatatgagctttag 51636
29 L S F T S T V X M T L K R N F F M A N M S L *
dp1ORF147
55207 atgtatctgtcaaagaagcgaataaggttattgaagatttctccacgagttccctaaaagtgccagactatatcatattcgctc
1 M Y L S K K R I R L K I S S S P S S L K W Q T I S Y S F
55291 aacagcaggcgcaggaacttgggatattgttcaaacagctaccggctcgaagaagaaggcttcttgatga 55359
29 N S R R R R T W D M F K Q L P V E E E G F L I *
dp1ORF148
28636 gtgtttcgttcaagaccattcgagtagggcgaacacctgtacgattttcgatgtcatccattgctgctaaatgtcagcgata
1 V F R F K T I R V G R T P V R P S M S S I A A K M -S A I
28552 gggctcactttcagctgggttagtccatttcttagtgactgcatattgttgcttagcatccattgttag 28484
29 G S L S A G L V H P L V T A Y C C L A S M L *
dp1ORF149
26474 atgccattgaacttttcgagcataaggattaaccttgcacctgctcactccagctgtggcggaatggctaaggttagttcg
1 M P L N F S S I R I N L A P L S H S S C G G M A N G S S
26390 agcaagtcgaaggcattgtattcgagattttgatatttatgagcagcaggtttccctag 26331

389

29 S K S K G I V F E I L I F M S S R F P *
dp1ORF150
15185 gtggctcctttacagcaagaaggaagtttattcgacctcttgacggttagatgcttcgcaagttcgacgattcgtttgttcat
1 V V L Y S K K E V Y S T S C T L I V F A K F D D S F V H
15101 ttgctttcgtgattgttcatgcaataggctcctcgatatttaagtttcacaagttcgctcgacgtg 15033
29 L L S L I V H A I G S S Y L I V S Q V A S T *
dp1ORF151
28027 atgattatatcaacgcaggaggagattgctagctacattcaagcacttccttcaaacgctcttcaatacccttggaccaactcttt
1 M I I S T Q G R L L A T F K H F L Q T L F N T L D Q L F
28111 tcctaatgctcaacaacaggagacacatttcatggtcgaaggtgcaataattgcccagtaa 28176
29 S L M L N K Q G Q T F H G S R V Q I I C Q *
dp1ORF152
42235 atgtgcataaaggacttatcgacaagaggctactattgcagctacttctgaaggatttagaccgaaagtttcaatgtatcttc
1 M C I K D L S T K R L L L Q Y F L K D L D R K F Q C I F
42319 aggtcttcaataactcatatggaaatgccattctatgtatatacactgacggaagacttgtggtga 42384
29 R L S I T H M E M P F Y V Y T L T E D L W *
dp1ORF153
22307 atgggtggcaaaagggtcaccttttcgaactttcgatctgctcatagcagacggttccattcgttcaggaaaaacagtatcgat
1 M V D K G L T F S N F R Y R H S R R P H S F R K N S I D
22391 ggctcttccatttcccttgggcatgacggaattcaacggacaaaacttggcattcgtggtga 22456
29 G S F I F P L G H D G I Q R T K L C H L W *
dp1ORF154
18446 gtgacaataggctttaagaactgcaaaaaaactggggcgctctgcacgcgcaacctggagctccttaacagctcatccaaggctg
1 V T I G F K N C K K T W G V C T R N L E L L N S H P R L
18530 aggtttcttacaacaactcctaatccttcaaaatagctcttgcgggtgcaatagtcctaa 18592
29 R F L T N N P N S F K I A L V R V N S A *
dp1ORF155
13512 atgaatacagacctgagcaacttacaatgggacatggtgcaaaatcctaatcttcttcaacgtttcattcaactcacgcccag
1 M N T T L S N L Q W D M V Q N L I S F P N V S F N S R Q
13596 ttgaagctcaagcaattttctggcatatgggagcctatgatattagctcttatgcaaatgtga 13658
29 L K L K Q F S G I W E P M I L V L M Q I *
dp1ORF156
18777 atgctagtagtctccatttctgtgtgcttctgttttagctctgttcagttcagctgcttctcgcatgcaatagtttcgagaat
1 M L V S P F L L V L L F S S V Q F S C P S R C N S F E N
18861 atgcctgttcataggctcacaatattccgcaaaagatttgcagttatggtggcgtaattaa 18923
29 M P V H R L T I F R Q R F A S Y G G V N *
dp1ORF157
13281 gtgcttgcgtgacttgagaagaattggtatcattttcgagccaatccataaggttctcgataccgtcacgattgattgtttct
1 V L A G L E K K L V S F S S Q S I R P S I P S R L I V S
13197 gttactgctttcttgaagcgttttttaagctgtcattatagacccttctcattttctataa 13135
29 V T A F L K R F L K S V I L D P F H F L *
dp1ORF158
40727 gtgaacgcgcttattagggctcaaacgaagcccaacggacattgtcttctgcccgtcactattgtgaggaaacagtcacttctcc
1 V N A V I R V K R S P N G H C L C P V T I V R N S H F S
40643 acttgcgagcgttacctcttccgagcgtgtcgttagctctgggtgactgctatgaacactga 40581
29 T C E R Y L P A G R V V V W V T A M N T *
dp1ORF159
30371 atgatttggctcgcgttaccacagcagcttctcctttgagtttctgtcgagcattccctgtacggtctgtccaaatagcatgc
1 M T E V A L T Q A A S P L S F C R A F P V R S V Q I A C
30287 gttcttgcgtattcttccatttagtagcagcagcttccgagactgttatgacagcagctga 30225
29 V F A Y S S I L V A A T S Q T V M T A T *
dp1ORF160
41324 atgggttacagacacgcgaggaaaacaatcgaacgtccaagacgtatctatcaatgttatagaatactatggaccgtctatcaa
1 M G Y R H A R K T I E R P R R I Y Q C Y R I L W T V Y Q
41408 tttctccgttcaacgtactcgtcaaaatcctgcaattatccaagctcttcgaaatgctaa 41467
29 F L R S T Y S S K S C N Y P S S S K C *
dp1ORF161
52175 atgcaaaaagggtttaaatgcttattctcgacatgacattgaaagcattgcattcgagactatttcaaatgttggcaacgttca
1 M Q K G L N A Y L D M T L K A L H S R L F Q N V W Q R S
52259 aatcaaaccaaggggccaagttttcaactaccttacaagactcttcaagaatagaatag 52318
29 N Q T K G P S F Q L T L Q D S S R I E *
dp1ORF162
13020 atgacagaagttgcggtaaatagccccgaaaagggtgagagtagttatggtcgggaatattgaatttctcgaatatttaaaaagg
1 M T E V A V N S P Q K V R V V M V G N I E F L E Y L K R
13104 aagtacggaacagaaacttccatcagttatattatagaaaatgaaaggggtctaataatga 13163
29 K Y G T E T S I S Y I I E N E R G L I *
dp1ORF163
40224 gtgaccgaatttctatgttctccgcagggaatgaagttatgtaccttacgcaagggttcattcacatcgataacgggactcttta
1 V T E F L C S P Q G M K L C T L R K G S F T S I T G S L
40308 cccaattcattcaagtcggcgattttagaacggaacaaactcgttttaattccagacatga 40367
29 P N P F K S A D L E R N N T R L I Q T *
dp1ORF164
6696 atgtactcttggagaacttctgctcctaaatgttccagcttccgcccattgcaattagggttagaactcgtgttatctataatagac
1 M Y S W R T S C L N V P A S P I A I R L E S A L S I I D
6612 tcaccgattcttccgaaatacatttttcgaatacattccaccaaccccgctgggcttataa 6553
29 S P I L S K Y I F R I H P P T P L G L *

dp1ORF165
50504 atgagtgaaagctgggtcaatccccaccacagatgggtctatatatttagatatcatgctatctaaaattgcaggggtaagggtctctt
1 M S E S W S I P T T D G L Y L D I M L S K I A G V R F F
50420 cctccaatcataaagggtgctgactaccacaagggaattttcagcctcagtcattgcttga 50361
29 P P I I K G V T T T R E F S A S V I A *

dp1ORF166
23519 gtggtcatgctcttttaactctatcttctccggttggctcgctttactgtcccagctgtaagcatagattcatcaatgct
1 V V M L P N D S I F S R L A R F T V P A V S I V F I N V
23435 gtgcgtgttgcagggcagtgtaaatctattctcagccaagagttcagcgtgaaatga 23376
29 V R V A R V E C K S I L S Q E F S V K *

dp1ORF167
1008 atgcttattcgggtggagcttcttacgtcgtatatgggtgctcacgcagacgagtcggggtggttaccctgattgcactc
1 M L I R L E L L T S Y M V L T Q T M R L E V L T L I A L
1092 ctgagttctataattcaatgtcaatgcaatggaatggaactggaggcaaggtaa 1148
29 L S S I I Q C Q M Q W N M E L E A R *

dp1ORF168
54345 atgagactttttccagggttatattcttcacattgttcagttctcggagtcgaagtattgttcttgaattcatagagttcgaaag
1 M R L F P G Y I L H I V Q F L E S S I V L E I H R V R K
54261 ttgcgaagggtcataggccgcatacatataggcaactcaggagggaattaaactaa 54205
29 P A K G H R P H T Y R Q H Q E E L N *

dp1ORF169
45954 atgaacacagcatcggaagagttcaatgttagtgataagggaagattcgtcgtggccaccaagcaagttctctgccgttta
1 M N T A S R R V S M L V I R K N S S W P P S K S S A R L
45870 gaaactcgtcaatcactaatttccatcttttagtgactcgaacttcttaaatatga 45814
29 E T P S I T N F P S L V T R L P K I *

dp1ORF170
27600 atgatgattgttcttctgtcctcctgccgtttgttgagcagcagcaagttgcttaccaaaagagccgatttcacgaggttcgggaa
1 M M I V L V L L P F V E Q Q Q V A Y Q K S R F H E V R E
27516 caccaccacgcacagcagctggatttcttaatttccagtcggcgctggcgacttag 27460
29 H H H R H D L D F L N F Q S R L A T *

dp1ORF171
47678 atgtcattttcttctcatgtactcttttagagcatcagcaagacttttgactgtttctccatgtcgccttctgtagcatttaat
1 M S P S P M Y S F R A S R R L L T C F S M S P L V A F N
47594 tcacgggttcttcaattgcagcagtgaaactgttttctcttcaatttcatattaa 47538
29 S P A S S I A A M N C F S S S N F I *

dp1ORF172
10462 atgttttcgaacattttctacccattattagaagcagcatcaatttcaataggagagccaagtcctttgttcacatccttcgctg
1 M F R T P S T P L L E A A S I S I G E P S P L F T S F A
10378 aaaattcgagcagtagtggttttaccagttccagcgccaccacagaatagatag 10325
29 K I R A V V V L P V P A P P Q N R *

dp1ORF173
32160 atgacattagacatttcttctgtctgtacgaaagggttcagcttgagtcacttcacgctacattgcactgaagattgtcataag
1 M T L D I S F V C T K G F S L S H F T V H C T E D C H K
32076 ttgctcatctgcataatactcgccgacttcagcgtgaagtaggtctaccattga 32023
29 L L I C H I L A D F S V S R L Y H *

dp1ORF174
29766 atgtcccatcagcccttttcattaagattgtcgaaccagcgttcgacttttccatcagtttcaagctgttcttcttatattggt
1 M S H Q P P S L R L S N Q R S T F H Q F Q A V L A Y I G
29682 cataatagaattgcgccattgtttccagtagtctgcgtcaccttttagactga 29629
29 H N R I A P F V S S S L R H L L D *

dp1ORF175
15648 atgcgcgtgatgtcatggcagataggcgaggataaaagagtgatgaatagaacgcgcagagcttacgagagcgccaaatacaag
1 M R V M S W Q I G E D K E C R I E R R R A Y E S A K Y K
15564 ggcgacggtactacgggtggtcctcttgccttaccgttaaccaataaaccattga 15511
29 G D G T T V V L L L T C N Q I N H *

dp1ORF176
43031 gtgataaagacggtaacgttgaaattttctagttccgtcttgaatgacgtcattttggtgattgattgctactgctggttggct
1 V I K T V T L N F S S S V L N D V I L V I D C Y C R L V
42947 aatcccgctcgacctgctgtttaagagtgctaaagagtgtagagatattccttaa 42894
29 N P V D L L F K S A K S C R D I L *

dp1ORF177
19937 atgaacctaaacagttcgagacttctcaagctgttgggaaagaagcaggtcgaatattttggtgggaacgtgaacttggtcata
1 M N L N S S R L L K L L G K K Q V E Y F G G N V N L V I
19853 ttctcgcgactaatttttagtgcttttgtattaatcagcgtgatatgcgcttga 19800
29 P S R L I L G A F V L I S V I C A *

dp1ORF178
11924 atgacaactgtcgaccaatttaaaagacagttgaggaaggttaggctcaatttttccctcatcagtttctttaaatttgagc
1 M T T V D Q F K R Q L R K S L G S I F P S S V - S L - N - L S
11840 caattagtaacctttagcgaattgttagcacttgcctcccatattaagtcataa 11787
29 Q L V T F S E L L A L A S H I K S *

dp1ORF179
56058 atgggttagggttattccttaccctcgttgatttgccttatgcaaaacctaccacaatcgcttgcgtgggttcaggagttgcatt
1 M G R V I P Y L V D L L Y A K P T T I A C R G F R S C I
56142 ttggataagtcaaaagcaagtgcttttatattcgacaagctctcgaataa 56192

391

29 L D K S K S K C L Y I R Q A L E *
dp1ORF180
41176 atgttcgacatgatttggaggaagtgttccctgttataatctgtcgaaccgctgaagtcgtgtctactaaagaaatgcccgaa
1 M F D M I W R K L F P V K I C R T A E V V S T K E M P E
41092 aaagtaggacgtactgaatcggggatgttgaacctccatccgtttgaatag 41042
29 K V G R T E S G M L N L H P F E *
dp1ORF181
13126 atggaagtttctgttccgtacttcccttttataatctcgagaaattcaatattcccgaccataactactctcaccttttgcggg
1 M E V S V P Y F L F K Y S R N S I F P T I T T L T F C G
13042 ctatttaccgcaacttctgtcataggtgtccttctgttatactgtaa 12992
29 L F T A T S V I G C P P L L I L *
dp1ORF182
45369 gtgcttgcctcatgtttcaataaagggttcgacctcgcttagcttgcgaacgtgctataacgatttcaatcatagcgaagaaa
1 V L A H V S I N R V R P R L A P E R A I T I S I I A K K
45285 ggtgagaagcttcaatccattcgggtgtcgaatcttcttccctga 45235
29 G E K L Q S I P L R C Q Y L L P *
dp1ORF183
13896 gtgattccagcttttgggttttcttccgctcttcaacttttcttcccttaggcgaggttttcttacgagttgaactcttaggt
1 V I P A F G F S S A S S T F S S L G A G F L R V E L L G
13812 ttttctcaactacttcttcaacctcagctcttgttcaactggacctga 13762
29 F S S T T S S T S A S C S T G P *
dp1ORF184
53330 gtgaacttgcgctcaaccacgtcaaacatttggcttctcgtcgaaggtctaaaattagagttccaagaagtcgctcttttctgga
1 V N L P S T T S N I W S S S R S K I R V P R S S L F S G
53246 aaatcttcaagagtagcactgtcttccggacgtcttgaaggaattcataa 53196
29 K S S R V A L S S G R S G R N S *

dp1ORF185
22522 atgaaattcgagatgttcgaatgaaatctactattattagacactttagaatggcgaagaattgtcaactacttctata
1 M K F E M F E M K I Y L L L D T L E M A K K L S T T S I
22606 tatttggaggaaaagatgagtcgagtcgaagacctatacaggggttaa 22653
29 Y L E E K M S R V K T L Y R G *
dp1ORF186
21272 atgctcgaaaaactcaaccggttcgaaaacctcaatcttgcgaaggctgaaccattcgaaaagttcaaaagttcgaaaaactc
1 M L E K L N R F E N L N P S K S R T I R K V Q K F E K L
21356 aaccattcgagtaggaattaaaggacataccagttcaacttttttag 21403
29 N H S R V G I K D I P V Q P F *
dp1ORF187
34415 atggtcttgttcaatctcttctactatcattcaagcagctgttcaattatcactgctttattcaatggtcttgttcaggcac
1 M L F N L F L L S P K L S L L Y S M V L F R H
34499 ttcctacgcttattcaagcaggtcttcaattttgtcagctctcataa 34546
29 F L R L F K Q V F K F C Q L S *
dp1ORF188
35609 atgttcgtaaaagcagcggttcgctcagtggaacttgttcaatcaggaagtgacaaccttaaccaacctcagtcacaatcta
1 M F V K Q P V R L E W T C S I Q E V T T L T N L S H N L
35693 aaaacaatcaagcgagcaaacggtgttcaacttgcgaacaatcgtag 35740
29 K T I K A S K P L S T L E Q S *
dp1ORF189
42587 atgcaaacgcagtatcaaccgtctctgaaactcttcatgacctgagacttgtatgctgcgaaccgtcgagaacttcgagctgacg
1 M Q T Q Y Q P S L K L F M T Q T C M L R T V E N F E L T
42671 agcaaaaacttcgcgaactcgttacgcaatcgaagatgaattcttag 42718
29 S K N P A K L V T Q S K M K F *
dp1ORF190
39786 atgtattcactcaaagttgttctcagtggttcaatcatattaaatcgaacttggttaatatcttactccttttagtgaagcag
1 M Y S L K V V Q C G S I I L K S N L V I S L L L L V K Q
39870 aggaagaccttaaatatcgaattgactcaaaagccgatcaaaagctaa 39917
29 R K T L N I E L T Q K P I K S *
dp1ORF191
40996 atgtccattgttccggaacttgatttaggttaagtaccttgcgaagtcagtcgagcgttaaggatacgttagtagtattggttc
1 M S I V P E L D L G K Y L A K S S D G V K D T L V V W F
40912 ttacctaaatctatccagtcgctaccgaaaactcgggtaccaacttga 40865
29 L P K S I Q S L P K T R Y Q T *
dp1ORF192
2920 atggctcgagtcgaatgtttttcgagatgaagtttagggcttctcgtatccctacggatgttcagcgagtgctttaacaaa
1 M V D V E C F F E M K F R V F S I P Y G M F S E C F N K
2836 acggaatggagtattcttgaacctgtcagttctcgtcctcgcctaa 2789
29 T E W S I L Q P V T F C V L A *
dp1ORF193
42456 atgatttcagctcaaattaaatcgaatgagacattgtctaaatttaaccaagaattatctacattcagatttcacacaaagt
1 M I S A Q I K Y E M R H C L N L T K N Y L H S L S P Q V
42372 ttcctcagtggtatatacatagaatggcatttccatagatttga 42325
29 F R Q C I Y I E W H F H M S Y *
dp1ORF194
40284 atgaaccttgcgtgaaggtacataacttctcctcgcggagaacatagaaattcgggtcacttgataccttaattggtagagcta
1 M N P C V R Y I T S F P A E N I E I R S L D T L M V E L

392

40200 ccgtcgcttcttaccgataaattagaccttcattagaagagctcatgtaa 40153
29 P S F L P I I R P S L E E L M *

dp1ORF195
42584 atgttcacaatcggtgttttgacaagttttctttcagctccttgcctaatagtgaaactctgccacaatttggcgcgattttgta
1 M P T I V V L T S F F S A P C P I V N S A T I W R D F V
42500 aggttcaacatagttctcacctcctttctaaaaaatattataacatga 42453
29 R F N I V L T S F L K N I I T *

dp1ORF196
11273 atggtagatttaacaagtcctgtccaatcatgtcactcctccttgcctcatcaaaagaagtttggtttcaattatcggttttagc
1 M V D L T S P C P I M S L L L A H Q K K F G F N Y R F S
11189 attaggctcccatTTaacaactccagcaagttcatttcttcttag 11142
29 I R L P F N N S S K F I H F P *

dp1ORF197
7484 atgaaaagattatatgggtatccaatttcaagccttgaaaaaattaaacggctctggagttaaaagcgtcaacccaaacttcatcg
1 M K R L Y G I Q F Q A L K K L N G L E L K A S T Q T S S
7568 agcgagggtatgaagttcttacaagaagcgtcgaactagatga 7612
29 M Q G M K F L T R S V E L D *

dp1ORF198
24119 atgccgctcaacaaattgaagtccttcttattcaatgcctcagttcacctatacagttgacctagaacccctccagcttgc
1 M P L N K L T S S F I Q C L S S P I Q L T L E T L P A C
24203 tcttctgtgacattgtttatcaggacgagcgtacaaaaggaaatga 24247
29 F L L T L F I R T S V Q K E *

dp1ORF199
15742 gtggctctgaattaggtgtacttttctcccaactgcttagcaactgccttctctttagcactagctctgcgcgtggga
1 A P E L G C T F P P N C L A T A F S C L A L A L R V G
15658 attggtttgtatgcgcgtgatgtcatggcagatagcgaggataa 15614
29 I G L Y A R D V M A D R R G *

dp1ORF200
47843 atgacaggcttgattcgcataagccctgaagtttttccacacatttcttccgtctcgtcactaatttttcgataatt
1 M T G L Y S I S P E S F S H I S S V S A S S T N F S I I
47759 tctttcaagcgttcttctgctcattagttgagcgtctgtcgtgtag 47715
29 S F K R S S S I V E R S V V *

dp1ORF201
38569 atgggcttcacaagttccttcttatacaaaaggtcaatatcttggactcgaactatttggacctataccgattcaactaccga
1 M G F T S S F P N Q R S I S L D S N Y L D L Y R P N Y R
38653 aacggggtatcaaaaaacctacattccaaaagacgggaatga 38694
29 N G L S K N L H S K R R E *

dp1ORF202
44483 gtggggcgctttatTTttataaaaatttttatacaaaatgttgacaacattcactcattatcgtataatacaattatataaaata
1 V G R L F F I K I F Y K M L D N I H S L S Y N T I I K I
44567 aataaagccgaaaggcgaggagacattatgtcaaaatga 44608
29 N K A E R R G G H Y V K N *

dp1ORF203
22781 gtgattaggattggcgggttacaagagaaccacattttcgaacctgttacggaacagcgccctgtcgttgggtgacaaacga
1 V I R I G R V T R E P H F R T C Y G T A P C R L V D K R
22697 ttcaggcatcagtgccacctcatcacagaagatacctgctaa 22656
29 F R H Q C H L I T E D T C *

dp1ORF204
1471 atgaccacgggttcgagtcagggtatgtgtgactttatcacgtcaagaaaatcgaggtacattcattgacagacttgacc
1 M T T V R V K G W L L T F I T S R K S Q V H S L T D L T
1555 acgctgttcttcttcaagggaatgaaccaatcgcttttag 1593
29 T L F P P K G M N Q S L *

dp1ORF205
8524 gtgacactgatgaattgttctcagtttggtatgctactcgtgacgagatattcttctacgaccaagaattgccaatttagaa
1 V T L M N G S Q F G M L L V T Q I S S T T K E L P N L E
8608 ttcaggaaaagcaacctgctatcaagttcaatttctgtag 8646
29 F R K S N L L S S S I S *

dp1ORF206
19855 atgaccaaggttcacgttccccacaaaatattcgacctgcttcttccccacagcttgagaagttctgaactgttttaggttcac
1 M T K F T F P P K Y S T C F P P N S L R S L E L F R F I
19939 aaattgttcaacttgagcaagtgcatatttcttttag 19977
29 K L F N L S K C D I I L *

dp1ORF207
27502 gtgtcggtggtggtgttccgaacctcgtgaaatcggtcttcttggtaagcaacttgctgctgctcaacaaacggcaggagcac
1 V S V V V F P N L V K S A L L V S N L L L L N K R Q E H
27586 aagaacaatcatcattctttaaataataggaggaaactaa 27624
29 K N N H H S L N N R R N *

dp1ORF208
47279 atgtttggtatgaagcaaaagacttctgtaagaaaaataacattcacttcccgtttgttcttcttgaacctagaacagacctg
1 M P G M K Q K T S L K K I T F T S R L F F L N L E Q T L
47363 accatcgtggttctcagattctgggatgacgaaggcgtga 47401
29 T I V V L D S G M T K A *

dp1ORF209
29784 atgttaagaatcaagttcgttagagccattgaaccgctcctactaaaatcaaggtacttcgaaactcttgggtcagtgatggat

393

1 M L R I K F V E P L K P L L L K S R Y F E T L G S V M D
29868 atggaggaaagaaaaaggataaaagcgaatgaagtcgtag 29906
29 M E E R K R I K R M K S *
dp10RF210
53077 atgtttcaacttttcccgatcatggttgtaagtggaagaaatagttttcaatacagagggaatccgttttggcataatggac
1 M F Q L F P Y H G C K V E E I V F Q Y E G I R F G I M D
52993 aattatcaggatggactgtttcccgctcttcgccaatag 52955
29 N Y Q D G L F P R L R Q *

dp10RF211
20959 gtgctcgacttttatgtcgccctaattttgtttttacttacggactatgggattttaggtattttcagggcgctttttat
1 V L D F Y V A P N F C F Y L R T M G F V G I F R A L F Y
20875 ttacttattaagtcctttctatattagattgtttataa 20837
29 L L I K S F S I L D C L *
dp10RF212
52983 atggactgtttcccgctcttcgccaatagcattgcaattgatatagcgtcgacgacgctcaacgtctgcttcggtgactacgaa
1 M D C F P V F A N S I A I D I A S T T V N V C F V D Y E
52899 ataatccatgtcttcgcttccgggtcatcatacaatag 52861
29 I I H V F A F R V I I Q *
dp10RF213
30291 atgcgtctttgctattcttccatcttagtagcagcagacttcgagactgttatgacagcagacttgaaactgtttcgataccg
1 M R L C V F F H L S S S D F A D C Y D S D L K L V S I P
30207 ttacaggttactaacaattcttcaggcttccatactaa 30169
29 F T V T N K F F R L P Y *
dp10RF214
24273 atgatccaaagtgtttttcagtcgtctattcctttgtacgtcgtcctgataaacaatgtcaacagaaaagcaagctggaag
1 M M P K L P F S A H S F C T L V L I N N V N R K Q A G R
24189 gtttctagggtcaactgtataggtgaactgaggcattga 24151
29 V S R V N C I G E L R H *
dp10RF215
35822 atgttaccaaacctgatagagtttcttacttctattatacaatcctctcgacagtttgtcaacgtcgtcattgtttcgaact
1 M L P N P D R V S L L L A L Y N P L D S L S T S S L F R T
35738 acgattgttccaatgttgacaacggtttgctcgcttga 35700
29 T I V P M L T T V C S P *
dp10RF216
32849 atggcctcggagctcggcgccacatctctccagatacggcagccaggtcaagtaccctggcatagcgtccatgatttcatt
1 M A S E L A A T S P P D T A A R S S T P G I A S M I S F
32765 acccgaaacgggtgaaagctagattttccataccttga 32727
29 T W K P A E A R F S I P *
dp10RF217
23443 atgaatactatgcttacagctgggacagtaaaagcgagccaaacgggagaagatagagtcattaaagagcatgaccactgcatgg
1 M N T M L T A G T V K R A K R E K I E S L K S M T T A W
23527 ataggaacagatagcctgtctcactgacgctctaa 23562
29 I G T D M P V S L T L *
dp10RF218
22029 atggaatgcttcgggaagaggttcgatatagactacaaattgagcgcgagaaaattacattgctccggggccaaatggcgacc
1 M E C P R K R F D I D Y K L S A R K L H C S G P K W A T
22113 aggaaattgaaggcgaggttaagataacttcgtag 22148
29 R K L K A R L K I T S *
dp10RF219
51388 atgattttatgctcgactttttcagttctccatttcttcgaaaacgcttcagggtgacgccttgccctaactacttcgtagat
1 M I L C S T F S V L P P L R N A S G L T P C L T T S L D
51304 gttccaaaattcctttcagccactggtttccatag 51269
29 V P K F L F S H W F P *
dp10RF220
6334 gtgaagttttctcggtgacggttgatacaatttcttcaagagtaagctgttaaggtggcaagtgaattctttcttcgaaact
1 V K F S S V T V D T I S F K S K L L R W Q V N S F F E T
6250 ttcttgccagcagatgcgtacatgatgtcttcataa 6215
29 F L P A D A Y M M S S *
dp10RF221
43507 atgactgctcaagttctatgtactatgctctccgctcagccggagcttcaagtgctggatgggagtcgaactgagtagatgc
1 M T A Q V L C T M L S A Q P E L Q V L D G Q S I L S T C
43591 acgcatggcttattgaaaacggttatgaactaa 43623
29 T H G L L K T V M N *
dp10RF222
13212 gtgacggatcagagaaccttatggattggctcgaaaatgataccaatttcttctcaagtcagcagcactcgataccatggaa
1 V T V S R T L W I G S K M I P I S S Q V Q Q A L D T M E
13296 gctatgaagggtggacttgcgagcactcattaa 13328
29 A M K V D L S S T H *
dp10RF223
14055 atgtggtggtacctgctggatgtgttcagatgtctactacttctacagtgaagtcgctgacgtttactacaagaaagatgtcg
1 M W W Y L L D M F E M S T T S T V K S L T F T T R K M S
14139 acgagcctgacgatgacagcgacattctttag 14171
29 T S L T M T A T F L *

dp1ORF224
13621 atgccagaaaattgcttgagcttcaactggcggtgagttgaatgaaacggtgagaaggaaattagattttgcaccatgtcccat
1 M P E N C L S F N W R E L N E T L K K E I R F C T M S H
13537 tgtaagttgctcagggtcgattcatatgctaa 13505
29 C K L L R V V F I C *
dp1ORF225
32991 gtgagcaacgggtgacgtatttcatcgctctgcccattgctcgttagtttctgcggtcgtatcagctgctgctcgagcaaatatc
1 V S N G C D V F H R L C H V A S F C V R I S C C S S K Y
32907 gtcagccacgtgacccgctgggttgcctctaa 32875
29 V S H V T R L V C L *
dp1ORF226
25191 gtggctgcgtacattagtttgaacttcagtgagcgcaagttgcttagcagaaagttcatcgctaggaattggatagtggtgttc
1 V A A Y I S L N P S E R K L L S R K F I A R N W I V V F
25107 gatagtcattgtcgtaagtggttgcataactga 25075
29 D S H C R K C L I T *
dp1ORF227
23115 atgactcaattagatggtagcgcttatgacgttttcgagaatccataaaggccgaaggttgttcattatagataccaaagtgcg
1 M T Q L D G S A Y D V S R I H K G R R L L H Y R Y Q S R
23031 ctgctacgaataacgggtcgaaattctatattga 22999
29 L L R I N G R I L Y *
dp1ORF228
10450 atgttcgaaacattattgaagattctagatacaagttatggacagcgagttcaaagtttcatcattgacgaggttcattatgc
1 M F E T L L K I L D T S L W T A S S K F T S L T R F I C
10534 tttcaacggagcatttaattgcgctgttga 10563
29 F Q P E H L M R C *
dp1ORF229
27634 atgtgcgagtttaagaaaactgattttaaaccactcgaagcattgtcgcaattcctgaccactacgttgcgttggctgctc
1 M C E L R K L I L I K P L E A L S Q F L T T T L L W L L
27718 aaattccagctaccgcagcaactcaagtag 27747
29 K F Q L P Q Q L K *
dp1ORF230
50723 gtgacgaaaaatccggcacttgaactatctgtcgttataaacccgatatggcgaagaccgaaaaatcatcgaatatatgtggg
1 V T K N P A Y L N Y L S L K T D M A K T E K S S N I C G
50807 acggtgaaactggaacctatactcttatag 50836
29 T L K L E P I L L *
dp1ORF231
31071 atgcgcgtgtcattgcgtttcacatcttcagttccctccgaggtcagcgcttcgagttctgctgtttctgcgctatctacgaca
1 M R V S L R F T S S V P S E V T A S S S A V S A V S T T
30987 aagttagctccgcgacttttggcaactga 30958
29 K L A P P T F G N *
dp1ORF232
29385 atgtcaattccattagctcttgcataattcaacgagctcaggaacggttttagccgcatactcttcgcgcatttgcctcaactcg
1 M S I P L A L A N S T S S G T V L A A Y S S R I C S T S
29301 tcaatttcttcaactgattcaattgtttga 29272
29 S I S S T D S I V *
dp1ORF233
52892 atgtcttcgccttccgggtcatcacaatagagtgacaattgcgctgtcaccgtggtcagcgagtgtaaaaaactcgattatta
1 M S S P S G S S Y N R V T I A L S P W S A S V K N S L L
52808 gaccctgagctaaatgttctgatttttga 52779
29 D P E L N V P D F *
dp1ORF234
36253 atgcttacgagtagacgagactcaactgttcgaaaggtttataagtttcaaccgcgttgggagcgatagcttacctaaccag
1 M L T S T A T Q L F E R F I S F N P L W E A I A Y L T Q
36337 gaagacctactcgacaatttagagtag 36363
29 E D L L D N L E *
dp1ORF235
32768 atgaaatcatggacgtatgccaggggtacttgacctggctgacctatctggaggagatgtggccgcgagctccgaggccatgg
1 M K S W T L C Q G Y L T W L P Y L E E M W P R A P R P W
32852 ctatgttcacttcgagcctttgattag 32878
29 L V H F E P L D *
dp1ORF236
37528 atgttcgtcgttttagatttagcaatatcagaggcttcagtggtggtgtagtaaacacgaaacatcaatgagatattcact
1 M F V A F R F S N I S R L H V A C S K P R N I N E I P T
37444 tccattgttgatagaagcaaacgttaa 37418
29 S I V D R S K R *
dp1ORF237
1678 gtgagagtcagggaaggaatcttgacatattctcagccgtagttctaaatccaaatagaactcgcttgggtgtaactgcattt
1 V R V Q V R N L D I P S A V V L N P N R T R L V S T A F
1594 gctaaagcgattgggttcattcccttga 1568
29 A K A I G S F P *
dp1ORF238
1301 atgcctttttgcggtcgatacaagttgcgcaagttccacaactttcagcgctcactttcataacatgaacgagtcagaaataag
1 M P F C G R Y K L R K F H N F Q R H F H N M N E S R N K

1217 gaacatctaaatcaattccccatttaa 1191
29 E H L N Q F P I *
dp1ORF239
26521 atgggtgaagtattttcctatcgagaatgtcctttcgaccatcctaagtggtaccacaaactgtatggtacgaaaactcac
1 M V K Y F L S K N V L S T I L M E C A T K L Y G T K T H
26605 tcgaagaaatcgctgatgagttga 26628
29 S K K S L M S *
dp1ORF240
41893 atgtttggaataagcgtgaaacagagttttacatggcgaagtaacaaatcaggagacaaccctacgggaactcgagggtgaatggg
1 M F G I S V K Q S L H G E V T N T R T T L R E L E V N G
41977 gactatttcaaaatttctggttag 42000
29 D Y F K I S G *
dp1ORF241
47020 gtgtctttccttaatatggagatagttttcattctattttaagcaggatcgaaggggttaccatttttagatttcattaggtt
1 V S F L N M E I V F I L F K Q D I E K V T N F R F H R L
46936 accatctacgatataatctgctaa 46913
29 T I Y D I I C *
dp1ORF242
41338 gtgtctgttaacccatgctcttcaggtagcggagccattaaagttcatcataccaatttgcgcgcgttttcgtgatagcttgg
1 V S V T H A L T V A E P L K F I I P N L P P F S L I A W
41254 tttttacctacgagctcagcgtga 41231
29 F L P T S S A *
dp1ORF243
51306 atgttccaaaattccttttcagccactgggtttccatagaaccctccatcggtttcgacctaatcacctcgagacgaattcagttc
1 M F Q N S F S A T G F H R T L H R P D L I H S R R I Q L
51222 gtccctgaagtgtagccgcaatga 51199
29 V L K C S R K *
dp1ORF244
27083 gtgaggtacaaaatgttgaccgtcgccgtcaatgaaaatttttagcatcgagttctttcgaagttttcgaaataatttccttcac
1 V R Y K M L T V A V N E N F S I E F F R S F R N N F L H
26999 ctggttgatagttggttcattctag 26976
29 L F D S W P I *
dp1ORF245
6278 gtggcaagtgaattctttctcgaaactttcttgcagcagatggtacatgatgtcttcataactgctagtagaagttttaat
1 V A S E F F L R N F L A S R C V H D V F I T A S R S F N
6194 tcgaagtcggtctttcaagaataa 6171
29 S K S V F Q E *
dp1ORF246
2831 atggagtatcttgcacccgtcacgttctggtcctcgcctaatagaccacaaaagtctttgaaagggtgctcagttattgtcca
1 M E Y L A T R H V L R P R L I D Q K V F E R L P Q Y C P
2747 aggttacaatttcacccggttaa 2724
29 R L Q F H P A *
dp1ORF247
29641 gtgacgcagactactggaacaaatggcgcaattctattatgaccaatataagcaagaacagcttgaactgatgaaagtcca
1 V T Q T T G N K W R N S I M T N I S K N S L K L M K S R
29725 acgctggttcgacaattcttaa 29745
29 T L V R Q S *
dp1ORF248
53560 gtgcaaacctcgttcttagcaagaagaacgatgctcagttacttgcctcaacggaaaaacaggaaagcctgcagttgaggttactt
1 V Q S L V L A R R T M L S Y L L N G K T G S L Q L R L L
53644 acatttcaggaaacgctctaa 53664
29 T F Q E T L *
dp1ORF249
2012 gtggatgcgactatcattgcaactggtgtgactcagcctttacctggaacggtactactgagccggaatatatcacaggcaag
1 V D A T I I A T G V T Q P L P G T V L L S R N I S Q A K
2096 aagctgctagtcgaatcttga 2116
29 K L L V E S *
dp1ORF250
23837 atgggcaaacatggaagattgacgaagactcagtcgactataaacctactcgagaatttcgaaactatattcgacaacttatca
1 M G K H G R L T K T Q S T I N L L E K F E T I F D N L S
23921 aaaagcaatcacgctttatga 23941
29 K S N H A L *
dp1ORF251
39205 atggaaataattagctcttaccgtctgcgcctgggttcccggtatcccttgagctccgtcattcccttccatttcgtccatgt
1 M E I I S L T V C A W L P G Y P L S S V I P L P F R P C
39121 ataggctgcagggtcttttga 39101
29 I G C R V F *
dp1ORF252
54771 gtgttgataggtcgaaactaattttgcatattttctatatttcaaaagtgttttgagatcgttatcaaaatgctcgacaa
1 V L Y R S K L I L H I F Y I S K V L L R Y R Y Q N A R Q
54687 tactttcgctgttctcttag 54667
29 Y F R L F L *
dp1ORF253
56255 atggttgctctataatagaaccgatgttgctagacaaagcatttgcaatcttcgagtcataatttattcgagagcttgcgaat

396

1 M V A S I I E P M L L D K A F A I F E S N L F E S L S N
56171 ataaagacacttgctttttga 56151
29 I K T L A F *
dp1ORF254
48479 atgaacctttcgcttaggttcaatcttttcgaacattttcatatttaacaaaactttcagctaaaaatcgacaaagttcaatg
1 M N L S L R F N L F R T F S Y L T K L S A K N R Q S S M
48395 ttcgactcaatgtttaataa 48375
29 F D S M F K *
dp1ORF255
9572 atgcttttggtcttctcgacgaatgactctactacattccctgcagggttttcgagcagtcagggtcaatgatgcaccgttttcgt
1 M L W S S R R M T L L H S L Q G F E Q Y G S M M H R F R
9488 caaggtagtcaccttttctaa 9468
29 Q G S H L F *
dp1ORF256
15289 atgaccttcagtcactaatgcggccgctgaaattggataccactatacatgggttcaccaacttcgagacaaagcagttgaaa
1 M T F Q S L M R P L K L D T T I H G F T N F E T K Q L K
15373 cacttgaagaaatttttag 15390
29 H L K K F *
dp1ORF257
28216 gtgaacgtgctggatttagcaacaagctactgagatggcattcttcctgtagtctatgcgacttggtgaaaagaccgtcaaa
1 V N V L D L A N K L L R W H S S V S L C D L V K K T V K
28300 acttgcgaatgctattga 28317
29 T C K C Y *
dp1ORF258
44023 atggaaattggtattggttcgaccgtgacggatatacattggctacgtcatggaaacggattggcgagtcaggtactacttcaatc
1 M E I G I G S T V T D T W L R H G N G L A S H G T T S I
44107 gcgatggttcaatggtaa 44124
29 A M V Q W *
dp1ORF259
4298 atgactcgactacgaagcataaagacaagtggatggaaagagtattcgaagtatttcgaaacagttctaatccagacgttaaga
1 M T R L R S I K T S G W K E Y S K L F R T V L I Q T L R
4382 ctccagcatttgggatga 4399
29 L T H L G *
dp1ORF260
24746 gtgacctacttctcctcaatcgccggtactggaggcaagcaagctcaagtcacttccatttcaggaaacttcaacttctctcag
1 V T L L P Q S A V L E A S K L K S L P F Q E T S T S F Q
24830 cggctgaatatttttag 24847
29 R L N I I *
dp1ORF261
288 atgaattcacttccctttgcccctaaaaacaggacagcctgacttcgcgaatgttttcattagttacattccaacgaaaagatgg
1 M N S L P F A L K Q D S L T S R M F S L V T P Q T K R W
372 ttgaatctaaatcattga 389
29 L N L N H *
dp1ORF262
9408 atgcctattcactccaggcggaagatgtggaagcatgcttgtgcagttcgacttaaattagaaaaaggtgactaccttgacg
1 M P I Q L Q A E R C G S M L V Q F D L N L E K V T T L T
9492 aaaacggtgcatcattga 9509
29 K T V H H *
dp1ORF263
27052 atgaaaatttttagcatcgagttctttcgaagtttttcgaaataatttcttcacctgtttgatagttggttcatctagacctttt
1 M K I L A S S S F E V F E I I S F T C L I V G S S R P P
26968 aacaagtcttctaattga 26951
29 N K S S N *
dp1ORF264
6139 gtgaatagtacaaggcggctctaatacgcctcaggatttctgctgtagggatagccgcacatcttccaaactcaattgagtcagc
1 V N S T R R S N T L R I S A V G I A A S S S N S I E S S
6055 tgtgaaacgtcttcataa 6038
29 C E T S S *
dp1ORF265
4801 gtgaataaagtcaagcgtttttgtataaaaagttcatttttttttaaaaaataagagcggaaggtcttatctaaaatagtc
1 V N K V K R F C I K S S F F F K K N K S E K L L S K I V
4717 gacgttgacgatttttaa 4700
29 D V D D F *
dp1ORF266
50220 atgcccgttcttccaagcagttgcaagcatttttatcaatagtcacgacttaccttgcagggtcgagccattatgacaatcaa
1 M P V L P S S C K H F I N S P R L T L S R S S H Y D N Q
50136 atcctcaccaggaagtaa 50119
29 I L T R K *
dp1ORF267
47367 atggtcaaggtctgttctaggttcaggaaagaacaaacgggaagtgatgttattttcttcagcgaagtccttttgcctcatacca
1 M V K V C S R F R K N K R E V N V I F F S E V F C F I P
47283 aacattaatcgtagatag 47266
29 N I N R R *
dp1ORF268

397

12621 atgtcaatttcggctcttgcttgacaatggattcaactactgatgcgccaaccccttttcaatcgcgacagcttgccaattca
1 M S I S V L C L T M D S T T D A S T F F N R D S L S N S
12537 ttgtcaattcttagagtaa 12520
29 L S I L E *
dp10RF269
53834 gtgaatagtatcgagtcctcagtttctacgtcaatagaacctattccgtcttcaatcattttgtctacatactgctcgagttt
1 V N S I E S I S F Y V N R T Y S V P N H F V Y I L L E F
53750 tgccttcctcagtgattaa 53733
29 C F L S D *
dp10RF270
50792 atgatttttcggctcttcgcatatcggtttttaacgacagatagttcaagtatgcgggatttttcgctcacgcttcatacgagata
1 M I F R S S P Y R F L T T D S S S M P D P S S R F I A I
50708 actctgctagcattttga 50691
29 T L L A P *
dp10RF271
19739 atgaggctgctttgctttatcttcggtaccgtattgaccgacttctactcgcgaaccttctacaagaattcatacctcaaag
1 M R L L C F I F V T V L T D F L L A N L P T R I H T S K
19655 gctttttgtcagccttag 19638
29 A F C Q P *
dp10RF272
1556 gtggtaagtcgtcaatgaatgtacctgcgattttcttgacgtgataaaagtcaacaaccatcccttgactcgaaccgtggtc
1 V V K S V N E C T C D F L D V I K V N N H P L T R T V V
1472 ataagttcgcctgctaa 1455
29 I S S A C *
dp10RF273
56256 atggatttcattaggactgagtcctcttggaattggaacgggttgcatatatagatattccgtcagccgtactaggccaagttct
1 M D F I R T E S S W N W N G C I Y R Y S V S R T R P S S
56340 agttcagtttatcttcgagtcattgcttcgagatatttgaaaaagtagtcaggaaaattcctgattatcttcgagtcattgc
29 S S V Y L A V N C F E I F E K V V R K I P D Y L A V N C
56424 ttcgagatatttgaaaaagtagtcaggaaaattcctgattattttttacaaaaacgcttga 56486
57 F E I P E K V V R K I P D Y F F Y K N A *

Table 31

Query= sid|114822|lan|dp1ORF001 Phage dp1 ORF|36698-40390|2
(1230 letters)

>gi|928828 (L44593) ORF1904; putative [Lactococcus lactis phage BK5-T]
Length = 1904

Score = 427 bits (1086), Expect = e-118
Identities = 226/475 (47%), Positives = 281/475 (58%), Gaps = 45/475 (9%)

Query: 395 AESGKYIGVLNTNKKPSELVDDFTWIRLEGPKG DAGLPAGPRDGVDPGKSGVGIAD 454
A+ YIG + P D+TW + +G+ G GA G+DGV GK GVGI
Sbjct: 820 ADYPSYIGQYTDIFIQYDSAKPSDYTWSLI---RGNDGKDGATGKDG---AGKDGVGIKT 873

Query: 455 TAITYAVSVSGTQEPENGWSEQVPELIKGRFLWTKTFWRYTDGSHETGYSVAYIGQDGN 514
T ITYA+S SGT +P GW+ QVY L+KG++LWTKI W YTD S ETGYSV YI +DGN+
Sbjct: 874 TVITYALSSSGTDKPNGTGWSQVPTLVKGGYQLWTKTVWYTDSSSETGYSVTYIAKDGN 933

Query: 515 GKDGIAKGKDGVGIAATEVMYASSPSATEAPAGGWSTQVPTVPGGQYLWTRTRWRYTDQTD 574
G DGIAGKDGVGI T + YA S T APA GW++QVP VP GQ+LWT+T W YTD T
Sbjct: 934 GNDGIAKGKDGVGIKTTITYAVGTSGTTAPASGWNSQVNPVAGQFLWTKTVWYTDNTS 993

Query: 575 EIGYSVSRMGEGQPGKDGAGR---DGIAGKNGIGLKSTSVSYGISPTDSAIP-GVMASQVP 630
E GYSV+ MG +G KGD G +GIAGK+G G+K+T+++Y SP + P G W++ VP
Sbjct: 994 ETGYSVAMMGVKGDKGDPGNGTNGIAGKDGKGIKATAITYQASPNGTAPTGTWSASVP 1053

Query: 631 SLIKGQYLWTRTIWYTDSTIETGYQKTYIPKDGNDGKNGIAGKDGVGIKSTTITYAGST 690
+ KG +LWTRTIWYTD+TTETGY Y+ +GN+G +G CKDG GIK+TTITYAGST
Sbjct: 1054 PVAKGSFLWTRTIWYTDNTTETGYAVAYMGNGNNGHDGFPKGKGTGIKTTITYAGST 1113

Query: 691 SGTVAPTSNTSAIPNVQPGFPLWTKTVWNYTDDTSETGYSVSKIGETXXXXXXXXXXXX 750
SGT P + WTS +P V G +LWTKTVW YTD+TSETGYSV+ +G
Sbjct: 1114 SGTTTPNNGWTSTVPTVAEGNYLWTKTVWYTDNTSETGYSVAMMG-----VKGDKGDP 1167

Query: 751 XXXXXXXXXXXXADGRS-QYTHLAFSNSPNEGEGFSHTDSGRAYVGGYQDFNPVHSDPAAYT 809
DG+ + T + + SPNG A G + P +K +T
Sbjct: 1168 GNGTNGIAGKDGKGIKATAITYQASPNGT-----TAPTGTWSASVPPVAKGSFLWT 1219

Query: 810 WTKN-----KGNDGAQGIPGKPGADCKTNYPHIAYASSADGS 846
T W GN+G G PGK G KT I YA S G+
Sbjct: 1220 RTIWTYDNTTETGYAVAYMGNGNNGHDGFPKGKGTGIKTT--TITYAGSTSGT 1272

Score = 396 bits (1007), Expect = e-109
Identities = 208/449 (46%), Positives = 260/449 (57%), Gaps = 42/449 (9%)

Query: 421 IRLEGPKG DAGLPAGPRDGVDPGKSGVGIADTAITYAVSVSGTQEPENGWSEQVPEL 480
+ + G KGD G PG +G +G+ GK G GI TAITY S +GT P WS VP +
Sbjct: 1155 VAMMGVKGDKG---DPGNGTNGIAGKDGKGIKATAITYQASPNGTAPTGTWSASVPPV 1211

Query: 481 IKGRFLWTKTFWRYTDGSHETGYSVAYIGQDGN SGKDGIAKGKDGVGIAATEVMYASSPSA 540
KG FLWT+T W YTD + ETGY+VAY+G +GN+G DG GKDG GI T + YA S S
Sbjct: 1212 AKGSFLWTRTIWYTDNTTETGYAVAYMGNGNNGHDGFPKGKGTGIKTTITYAGSTSG 1271

Query: 541 TEAPAGGWSTQVPTVPGGQYLWTRTRWRYTDQTD EIGYSVSRMGEGQPGKDGAGR---DGI 597
T P GW++ VPTV G YLWT+T W YTD T E GYSV+ MG +G KGD G +GI
Sbjct: 1272 TTPPNNGWTSTVPTVAEGNYLWTKTVWYTDNTSETGYSVAMMGVKGDKGDPGNGTNGI 1331

Query: 598 AGKNGIGLKSTSVSYGISPTDSAIP-GVMASQVPSLIKQYLWTRTIWYTDSTIETGYQ 656
AGK+G G+K+T+++Y SP + P G W++ VP + KG +LWTRTIWYTD+TTETGY
Sbjct: 1332 AGKDGKGIKATAITYQASPNGTAPTGTWSASVPPVAKGSFLWTRTIWYTDNTTETGYA 1391

Query: 657 KTYIPKDGNDGKNGIAGKDGVGIKSTTITYAGSTSGTVAPTSNNTSAIPNVQPGFPLWTK 716
Y+ +GN+G +G GKDG GIK+TTITYAGSTSGT P + WTS +P V G +LWTK
Sbjct: 1392 VAYMGNGNNGHDGFPKGKGTGIKTTITYAGSTSGTTPNNGWTSTVPTVAEGNYLWTK 1451

Query: 717 TVWNYTDDTSETGYSVSKIGSTXXXXXXXXXXXXXXXXXXXXXADGRS-QYTHLAFSNS 775
TVW YTD+TSETGYSV+ +G DG+ + T + + S
Sbjct: 1452 TVWYTDNTSETGYSVAMMG-----VKGDKGDPGNGTNGIAGKDGKGIKATAITYQAS 1505

399

Query: 776 PNGEGFSHTDSGRAYVQYQDFNPVHSDPAAYTWTKW-----KGND 817
 PNG A G + P +K +T T W GN+
 Sbjct: 1506 PNGT-----TAPTGTWSASVPPVAKGSFLWTRTIWYTDNTTETGYAVAYMGNGNN 1557

Query: 818 GAQGIKPGKPGADGKTNFYFHIAYASSADGS 846
 G G PGK G KT I YA S G+
 Sbjct: 1558 GHDGFPKDGDTGIKTT--TITYAGSTSGT 1584

Score = 384 bits (977), Expect = e-105
 Identities = 179/322 (55%), Positives = 222/322 (68%), Gaps = 7/322 (2%)

Query: 421 IRLEGPKGDAGLPGAPGRDGVDPGKSGVGIADTAITYAVSVSGTQEPENGWSEQVPEL 480
 + + G KGD G PG +G +G+ GK G GI TAITY S +GT P WS VP +
 Sbjct: 1311 VAMMGVKGDGK---DPGNNGTNGIAGKDGKGIKATAITYQASPNGTAPTGTWSASVPPV 1367

Query: 481 IKGRFLWTKTFWRYTDGSHETGYSVAYIGQDGNNGKDIAGKDGVGIAATEVMYASSPSA 540
 KG FLWT+T W YTD + ETGY+VAY+G +GN+G DG GKDG GI T + YA S S
 Sbjct: 1368 AKGSFLWTRTIWYTDNTTETGYAVAYMGNGNNGHDGFPKDGDTGIKTTTITYAGSTSG 1427

Query: 541 TEAPAGGWSTQVPTVPGGQYLWTRTRWRYTDQTDGIGYSVSRMGEQGPKGDAAGR---DGI 597
 T P GW++ VPTV G YLWT+T W YTD T E GYSV+ MG +G KGD G +GI
 Sbjct: 1428 TTPPNNGWTSTVPTVAEGNYLWTKTVWYTDNTSETGYSVAMMGVKGDGDPGNNGTNGI 1487

Query: 598 AGKNGIGLKSTSVSYGISPTDSAIP-GVNASQVPSLIKQYLWTRTIWYTDSTTETGYQ 656
 AGK+G G+K+T+++Y SP + P G W++ VP + KG +LWTRTIWYTD+TTETGY
 Sbjct: 1488 AGKDGKGIKATAITYQASPNGTAPTGTWSASVPPVAKGSFLWTRTIWYTDNTTETGYA 1547

Query: 657 KTYIPKDGNDGKNGIAGKDGVGIKSTTITYAGSTSGTVAPTSNWTSAIPNVQPGFPLWTK 716
 Y+ +GN+G +G GKDG GIK+TTITYAGSTSGT P + WTS +P V G +LWTK
 Sbjct: 1548 VAYMGNGNNGHDGFPKDGDTGIKTTTITYAGSTSGTTPPNNGWTSTVPTVAEGNYLWTK 1607

Query: 717 TVWNYTDDTSETGYSVSKIGET 738
 TVW YTD++ ETGYSV K+G T
 Sbjct: 1608 TVWAYTDNSFETGYSVKGKMGNT 1629

Score = 201 bits (507), Expect = 2e-50
 Identities = 121/297 (40%), Positives = 156/297 (51%), Gaps = 19/297 (6%)

Query: 421 IRLEGPKGDAGLPGAPGRDGVDPGKSGVGIADTAITYAVSVSGTQEPENGWSEQVPEL 480
 + + G KGD G PG +G +G+ GK G GI TAITY S +GT P WS VP +
 Sbjct: 1467 VAMMGVKGDGK---DPGNNGTNGIAGKDGKGIKATAITYQASPNGTAPTGTWSASVPPV 1523

Query: 481 IKGRFLWTKTFWRYTDGSHETGYSVAYIGQDGNNGKDIAGKDGVGIAATEVMYASSPSA 540
 KG FLWT+T W YTD + ETGY+VAY+G +GN+G DG GKDG GI T + YA S S
 Sbjct: 1524 AKGSFLWTRTIWYTDNTTETGYAVAYMGNGNNGHDGFPKDGDTGIKTTTITYAGSTSG 1583

Query: 541 TEAPAGGWSTQVPTVPGGQYLWTRTRWRYTDQTDGIGYSVSRMGEQGPKGDAGRDGIAGK 600
 T P GW++ VPTV G YLWT+T W YTD + E GYSV +MG GP AG +G GK
 Sbjct: 1584 TTPPNNGWTSTVPTVAEGNYLWTKTVWAYTDNSFETGYSVKGKMGNTGP---AGSNGNPGK 1640

Query: 601 NGIGLKSTSVSYGISPTDSAIPGVNASQVPSLIKQ-YLWTRTIWYTDSTTE--TGYQK 657
 + T+ G++ S + + ++ G +Y W W + G
 Sbjct: 1641 VVSDTEPTTKPKGLTWKYSGVVDMPLNGTKILAGTEYWNNGNNWALYEINAHNNGDNL 1700

Query: 658 TYIPKDGNDGK-NGLAGKDGVGIKSTTITYAGS-----TSGTVAPTSNWTSAIPNVQ 708
 + DGK I G +GV + T T GS +S + T N T A I N Q
 Sbjct: 1701 SVTNGTFKDGKIESIWGSGV---NGTTIEGSHLQIHSSDSTNTTEN-TLAIDNRQ 1753

Query= sid|114823|lan|dp1ORF002 Phage dp1 ORF|32386-35835|1
 (1149 letters)

>dbj|BAA31888| (AB009866) orf 15 [bacteriophage phi FVL]
 Length = 694

Score = 280 bits (709), Expect = 3e-74
 Identities = 157/465 (33%), Positives = 257/465 (54%), Gaps = 28/465 (6%)

Query: 40 QIGSALTGLGKGLTAVTLPLMGFAAASIKVGNFQAQMSRVQAIAGATAEELGRMKTA 99
 +IG+++ +G+ +T VT P++ A + K G EF M +V+A +GAT EE +K +A
 Sbjct: 151 EIGNSMKQVGRNNTMYVTAPVVAGFAVAAKGIEFDDSHRKVKATSGATGSEFEALKKKA 210

400

Query: 100 IDLGAKTAFSAKEAAQGMENLASAGFQVNEIMDAMPVGLDLXXXXXXXXXXXXXXXXXMASSL 159
 ++GA T FSA ++A+ + +A AG+ ++M+ + GV+DL + L
 Sbjct: 211 REMGATTKFSASDSAEALNYMALAGWDSKQMMGLSGVMDLAAASGEELGAVSDIVTDGL 270

Query: 160 RAFGLEANQAGHVADVFAAAAADTNAETSDMAEAMKYVAPVAHSMGLSLEETAASIGIMA 219
 AFGL+A +GH+ADV A+ ++ N + + EA KYVAPVA ++G ++E+T+ +IG+M+
 Sbjct: 271 TAFGLKAKDSGHLADVLAQTSSKANTDVRGLGEAFKYVAPVAGALGYTTIEDTSIAIGLMS 330

Query: 220 DAGIKGSQAGTTLRGALSRIAKPTKAMVKSMQELGVSYFYDANGNMIPLEQIAQLKTATA 279
 +AGIKG +AGT LR + ++ PT+AM M+ LG+S D+NG MIP+R+ + QL+
 Sbjct: 331 NAGIKGEKAGTALRTMTNLSSPTRAMGNEMERLGISITDSNGKMIPMRKLLDQLREKFK 390

Query: 280 GLTQEERNRHLVLTLYGQNSLSGMLALLDAGPEKLDKMTNALVNSDGAAKEMAETMQDNLA 339
 L+++++ T++G+ ++SG LA+++A E K+T ++ +S GA+K MA+TM+ L
 Sbjct: 391 HLSKDQQAASSAATIFGKEAMSGALAIINASDEDYQKLTKSIDSSTGASKRMADTMESGLG 450

Query: 340 SKIEQMGGAFESVAIVQOILEPALAKIVGAITKLEAFVNMSPIGQKMHVIFAGMVAAL 399
 K+ + E +A+ + +EPAL IV A +KV+ + Q VV F VA L
 Sbjct: 451 GKLRTRLRSQLEELALTIYDRIEPALKIIVSAFSAKVVTVVTKLPTSIQLAVVGFLFVAVL 510

Query: 400 GPLLLIAGM-----VMTTIVKLRIAIQFLGPAPMTGMTIAGVIAIF----- 441
 GPL+ + G+ MT + L I + P IA ++ +F
 Sbjct: 511 GPLVFMFGLFISVMGNAMTVLGPLLINVNKASGLFAFLRTKIASLVKLPILGVSISSLT 570

Query: 442 -----YALVAV---FMIAITKSERFRNFINS LAPAIKAGFGGA 476
 ALV + F AY +SE FRN +N + F A
 Sbjct: 571 LPITLIVGALVGIGIAFYQAYKRSETFRNIVNQAISGVANAFKAA 615

Query= sid|114824|lan|dp1ORF003 Phage dp1 ORF|53538-55877|3
 (779 letters)

>sp|P43741|DPO1_HABIN DNA POLYMERASE I (POL I) >gi|1074025|pir||E64098 DNA polymerase I
 (polA) homolog - Haemophilus influenzae (strain Rd KW20)
 >gi|1573871|U32767| DNA polymerase I (polA)
 [Haemophilus influenzae Rd]
 Length = 930

Score = 191 bits (481), Expect = 1e-47

Identities = 148/553 (26%), Positives = 262/553 (46%), Gaps = 60/553 (10%)

Query: 63 RLELITEBAKLEQYVDKMIEDGIGSIDVETDGLDITINDELAVCLYSPSQKGIYAFVNVH 122
 + E + +A L ++++K+ + ++D ETD LD + L G+ + + Y P+
 Sbjct: 333 KYETLLTQADLTRWIEKLNAAKLIADVDTETDSDLYMSANLVGISFALENGEAAVLPQLD 392

Query: 123 SNMTKMRINKQISPEFMKQMLQRIVDSGIPVVIHNSKFDMSIYWRIGVKMNEPANDTYL 182
 ++ + +K +L+ + I I N KFD +SI+ R G+++ +DT L
 Sbjct: 393 YLDAPKTEKSTALAAIKPILE---NPNHKGIGQNIKFD-ESIFARHGIELQGVFPTML 448

Query: 183 AANLLNENESHSLKSLHSKYVRNEENAEVAKFNDLPKGPFSLIPPDVAYMYAAYDPLQ 242
 + LN H++ L +Y+ +E A + + F+ IP + A YAA D T
 Sbjct: 449 LSYTLNSTGRHNMDDAKRYLGHETIAFESLAGKGSQLTFFNQIPLEQATEYAAEDADVT 508

Query: 243 FELYEPQEQYLTPGTEQCEYNLEKVSIVLHNIEMPLIKVLPDMEVYGVOLDQDLAEIR 302
 +L + L E Y +E+PL+ VL ME GV +D D L
 Sbjct: 509 MKLQALWKLQEEPTLVELYK-----TMELPLLHVLSRMERTGVLDSDALFMQS 559

Query: 303 EQPTANNBAEQEFQQLVSENPRIEELRQTNFQSYQKLEMDARGRTVTSISSPTQLAIL 362
 + + + E++ L + + + +S QL +
 Sbjct: 560 NEIASRLTALEKQAYALAGQ-----PFNLASTKQLQEI 592

Query: 363 FYDIMGLKSPERDKPRG---TGESIVEH--FDNDISXXXXXXXXXXXXXVSTYTT-LDQHL 416
 +D + L ++ P+G T E ++E + + + + STYT L Q +
 Sbjct: 593 LFDKLELPVLQKT-PKGAPSTNEEVLELSYSHLPKILVKHRLSKLXSTYTDKLPQMV 651

Query: 417 AKPDNRHITTFKQYGAKTGRMSENPNLQNIPIRGE-GAVVRQIPAASEGHIYIGSDYSQ 475
 R+HT++ Q TGR+SS +PNLQNIPIRGE+RQ F A EG+ I+ +DYSQ
 Sbjct: 652 NSQTGRVHTSYHQAVTATGRLSSDPNLQNIPIRNEEGRHIRQAFIAREGYSIVAADYSQ 711

Query: 476 QEPRLAELSGDESMRHAQEQLDLYSVIGSKLYGVVPYEECLEFPDGTINKEGKLRRNS 535
 E R +A LSGD+ + +A+ Q D++ + + + +GV +E T+++ R +
 Sbjct: 712 IELRIMAHLSGDQGLINAFSQGKDHRSTAABIFGVSLDE-----VTSEQ----RRN 759

Query: 536 VKSVLLGLMYGRGANSIAEQMNVSVKEANKVIEDFFTEPPKVADYIIFVQQQAQDLGYVQ 595
 K++ GL+YG A ++ Q+ +S +A K ++ +F +P V ++ + + + +A+ GYV+

401

Sbjct: 760 AKAINFGLIYGMSAFGLSRQLGISRADAQKYMPLYFQRYPSVQQFMTDIREKAKAQGYVE 819

Query: 596 TATGRRRLPDMS 608

T GRR LPD++

Sbjct: 820 TLFGRRLYLPDIN 832

Score = 46.9 bits (109), Expect = 5e-04

Identities = 34/123 (27%), Positives = 66/123 (53%), Gaps = 16/123 (13%)

Query: 663 EIKDQAKAEGI-----LIKDNGGKIADAQRQCLNSVIQGTAAADMTKYAMIKV 709

+I+++AKA+G + N + A+R +N+ +QGTAA+ K AMIK+

Sbjct: 807 DIREKAKAQGYVETLFGRRLLYLPDINSSNAMRRKGAERVAINAPMQGTAAADIIKRAMIKL 866

Query: 710 HNDASLKELGPHLMIPVHDELLGEVPIKNAKGAERLTEVMIEAAKDIISLPMKCDPSIV 769

++ + +++ VHDEL+ EV + S++ + M EAA +++ +P+ + +

Sbjct: 867 -DEVIRHDPDIEMIMQVHDELVFVEVRSEKVAFFREQIKQHM-EAAAEVLV-VPLIVEVGVG 923

Query: 770 ERW 772

+ W

Sbjct: 924 QNW 926

Query= sid|114825|lan|dp1ORF004 Phage dp1 ORF|40401-42440|3
(679 letters)

>emb|CAB07981| (293946) hypothetical protein [bacteriophage Dp-1]
Length = 532

Score = 1011 bits (2585), Expect = 0.0

Identities = 497/499 (99%), Positives = 498/499 (99%)

Query: 1 MTKFINSYGPLHLNLYVEQVSQDVTNNSRVSWRATVDRDGAYRTWTYGNISNLSVWNLG 60

MTKFINSYGPLHLNLYVEQVSQDVTNNSRVSWRATVDRDGAYRTWTYGNISNLSVWNLG

Sbjct: 1 MTKFINSYGPLHLNLYVEQVSQDVTNNSRVSWRATVDRDGAYRTWTYGNISNLSVWNLG 60

Query: 61 SSVHSSHPDYDTSGEEVTLASGEVTVPHNSDGTKTMSVWASFDPNNGVHGNITISTNYTL 120

SSVHSSHPDYDTSGEEVTLASGEVTVPHNSDGTKTMSVWASFDPNNGVHGNITISTNYTL

Sbjct: 61 SSVHSSHPDYDTSGEEVTLASGEVTVPHNSDGTKTMSVWASFDPNNGVHGNITISTNYTL 120

Query: 121 DSIPRSTQISSFEGNRNLGSLHTVIFNRKVNSTHQVWYRVFGSDWIDLGNHTTSVSFT 180

DSIPRSTQISSFEGNRNLGSLHTVIFNRKVNSTHQVWYRVFGSDWIDLGNHTTSVSFT

Sbjct: 121 DSIPRSTQISSFEGNRNLGSLHTVIFNRKVNSTHQVWYRVFGSDWIDLGNHTTSVSFT 180

Query: 181 PSLDLARYLPKSSSGTMDICIRTYNGTTQIGSDVYSNGWRFNIPDSVRPTFSGISLVDTT 240

PSLDLARYLPKSSSGTMDICIRTYNGTTQIGSDVYSNGWRFNIPDSVRPTFSGISLVDTT

Sbjct: 181 PSLDLARYLPKSSSGTMDICIRTYNGTTQIGSDVYSNGWRFNIPDSVRPTFSGISLVDTT 240

Query: 241 SAVRQILTGNFLQIMSNIQVNFNNSGAYGSTIQAFHAELVGKNQAINENGKLGMMNF 300

SAVRQILTGNFLQIMSNIQVNFNNSGAYGSTIQAFHAELVGKNQAINENGKLGMMNF

Sbjct: 241 SAVRQILTGNFLQIMSNIQVNFNNSGAYGSTIQAFHAELVGKNQAINENGKLGMMNF 300

Query: 301 NGSATVRANVTDRGKQSNVQDVSINVIEYGPSINFSVQRTQNPAAIQALRNAKVAPI 360

NGSATVRANVTDRGKQSNVQDVSINVIEYGPSINFSVQRTQNPAAIQALRNAKVAPI

Sbjct: 301 NGSATVRANVTDRGKQSNVQDVSINVIEYGPSINFSVQRTQNPAAIQALRNAKVAPI 360

Query: 361 TVGGQQKNIMQITFSVAPLNTTNFTEDRGSASGTFTTISLMTNSSANLAGNYGPKSYIV 420

TVGGQQKNIMQITFSVAPLNTTNFTEDRGSASGTFTTISL+TNSSANLAGNYGPKSYIV

Sbjct: 361 TVGGQQKNIMQITFSVAPLNTTNFTEDRGSASGTFTTISLMTNSSANLAGNYGPKSYIV 420

Query: 421 KAKIQDRFTSTEPSATVATESVVLNYDKDGRGLGVGKVEQKAGSIDAAGDIYAGGRQVQ 480

KAKIQDRFTSTEPSATV TSVVLNYDKDGRGLGVGKVEQKAGSIDAAGDIYAGGRQVQ

Sbjct: 421 KAKIQDRFTSTEPSATVPTESVVLNYDKDGRGLGVGKVEQKAGSIDAAGDIYAGGRQVQ 480

Query: 481 QFQLTDMNGALNRGQYNDV 499

QFQLTDMNGALNRGQYNDV

Sbjct: 481 QFQLTDMNGALNRGQYNDV 499

Query= sid|114827|lan|dp1ORF006 Phage dp1 ORF|45296-46987|2
(563 letters)

>gb|AAD18987| (AE001666) SWI/SNF family helicase_2 [Chlamydia pneumoniae]
Length = 1166

Score = 171 bits (429), Expect = 1e-41

Identities = 150/522 (28%), Positives = 254/522 (47%), Gaps = 55/522 (10%)

402

Query: 46 SSNNFE-LPYKYFNNVIDALDEWELHIFGELDKVDQYIDSRNRIASSNEQFSFKTTPF 104
S + FE LP + ++ + L E + I GE++ D QD + T

Sbjct: 659 SLDQFEALPVNF--SMSERLIEIQKQIRGEIEFDQD-----VPQIQATLSYQTEG 709

Query: 105 AHQVECFEYAEHPFCFLGDEQGLGKTKQAIDIAVSRKASFH--CLIVCCISGLKWNWA 162
H + E + H + L D+ GLGKT QAI IAV++ K C ++ C + L +NW

Sbjct: 710 VHWLE--RLRKMHLNGILADDMLGKTLQAI-IAVTQSKLEKSGCSLIVCPTSLVYNWK 766

Query: 163 KEVGIHNSAHILGSRVTKDGLVIDGV-SKRAEDLLGGHDEFFLITNIETLRDAVPIK 221
+E + E LVIDGV S+R + L D IT+ L+ V

Sbjct: 767 EEFKRFNPEFR-----TLVIDGVPSQRRKQLTALADRDVAITSYNLQKDV--- 812

Query: 222 YLNELTKSGEIGMVIIDEIHKCKNPSSKQASIQKLQSYKMGTLGTPLMNPIDVFNVM 281
EL KS V++DE H KN +++ S++ +QS +++ LTGTP+ N+ +++++

Sbjct: 813 ---ELYKSFPRPDYVVLDEAHKKNRTTRNAKSVKMIQSDHRLILTGTPIENSLEELWSLF 869

Query: 282 KWLGAEHHTLTQFKERYCIVDQFNQITGYR-----NLAELRELVDNMYLRRTKEEVL-DL 335
+L L +R+ V ++ + Y N+ L++ V+ ++LRR KE+VL DL

Sbjct: 870 DFLMPG---LLSSYDRF--VGKYIRTGNMGNKADNMVALKKVSPFILRMKEDVLKDL 924

Query: 336 PEKIRVTEYVDMNSKQSKIY-----KEVLTKLVQEIQKVLMPNPLAETIRLRQATGN 388
P + + + Q ++Y K+ L++LV++ ++ + LA RL+Q +

Sbjct: 925 PPVSEILYHCHLTESQKELYQSYAASAKQLSRLVKQEGFERIHIHVLATLRLKQICCH 984

Query: 389 PSILTTQDVK---SCKFERCIEIVEECIQQKSCVIFSNWEKVIEPLAKIL-SKTVKCNL 444
P+I + S K++ +++++ + G V+FS + K++ + K L S+ +

Sbjct: 985 PAIFAKDAPEFGDSAKYDMLDLSSLVDSGHKTVPVFSQYTKMLGIKKDLSESRGIPFVY 1044

Query: 445 VTGETADKPFNEIEFPMNRKASVILGTIGALGTGFTLTAKDTVIFLDSPWTRAEKQQAED 504
+ G T ++ + + +P V L ++ A GTG L ADTVI D W A ++QA D

Sbjct: 1045 LDGSTRNRLDLVNQFNEDPSLLVFLISLKAGGTGLNLVGADTVIHYDMWNPVAVENQATD 1104

Query: 505 RCHRIGAKSSVTIYTLVAKGTVDERIEDLIERKGELADYIVD 546
R HRIG SV+ Y LV T++E+I L RK L +++

Sbjct: 1105 RVHRIGQSRVSYSYKLVTLNTIEEKILTQNRKKSLLVKKVIN 1146

Query= sid|114828|lan|dp1ORF007 Phage dp1 ORF|22230-23621|3
(463 letters)

>gi|2444105 (U88974) ORF26 [Streptococcus thermophilus temperate bacteriophage
O1205]
Length = 411

Score = 88.9 bits (217), Expect = 7e-17
Identities = 80/315 (25%), Positives = 133/315 (41%), Gaps = 48/315 (15%)

Query: 139 QGVTLAGIFCDEVALMPESFVNQATGRCSVTGSKMWFSCNPNPNHYPKKNWIDKQVEKR 198
+G T G + +E +L E + RCS G+++ + NP NPNH+ ++I K + +

Sbjct: 121 RGFTAPGAYVNEASLANELVFKEIISRCSDGARVVWDSNPDNPNHNLNRDYIGKN-DGK 179

Query: 199 ILYLHFTMDNPNPLT----DSIKRREYKMYAGVFRKRFILGLWVTADGLVYSMFNEEQHV 254
I+ F +DDN L+ DSIK K G F R ILGLW A+G +Y+ ++ + HV

Sbjct: 180 IIDFSFKLDDNTFLSKRYIDSIAKATPK---GKPYDRDILGLWTVAGAIYADYDSKINV 236

Query: 255 KKLNIIEFDRLFVAGDFGIYNATTFGLYGFSKRHKRYHLIESYHSGREASEQLTEADVNS 314
E R P D+G + + + G ++L++ +E + + +A

Sbjct: 237 VDLPFMKRYFGGIDNGYTHYGSIVIVG-EGVDNNFYLVGVAAQFKEIDWWVEQA---- 291

Query: 315 NIQPSVLQKTTKEYANDLVDMIRGKQIEYIILDPSASAMIVELQKHPYIAR---KNIP 371
+K T Y N + + ++AR + I

Sbjct: 292 -----RKLGTIYGN-----IPFYADSARPEHVARFENEGFDI 323

Query: 372 IPARNVDTLGISPHAELLAENRFTLDPSNT-HDIDEYAYSWDSKASQTGEDRVIKZHDH 430
+ A V GI A+L E + + DE Y Y W ++ +D +KE D

Sbjct: 324 MNANKSVIAGIELIAKLFKEKKLYVKGFPVPRFFDEIYQYRWKENST---KDEPLKEFDD 380

Query: 431 CMDRNRVYACLTDALI 445
+D RYA +D +I

Sbjct: 381 VLDSVRYAIYSDYVI 395

Query= sid|114829|lan|dp1ORF008 Phage dp1 ORF|49624-50961|1
(445 letters)

>gb|AAD19901| (AF100420) DnaB replication fork helicase [Thermus aquaticus]

403

Length = 444

Score = 67.5 bits (162), Expect = 2e-10
 Identities = 69/248 (27%), Positives = 111/248 (43%), Gaps = 14/248 (5%)

Query: 147 GERLGISTGFEXXXXXXXXXXXXXXIVIMARPGQKS-WTIDKMLATAWKGHDVLLYS 205
 GE G+ TGF+ I I ARP GK+ + + A K G V +YS
 Sbjct: 178 GEVAGVRTGFKELDLGLTGLPGSLNI-IAARPAMGKTAFALTIAQNAALKEGVGVGIYS 236

Query: 206 GEMSEMQVGARIDTILSNVSINSITKGIWHDQFEKYEDHIQAMTEAENSLVVVTPFMIG 265
 EM Q+ R+ + + +N+ G D F + D ++EA + TP +
 Sbjct: 237 LEMPAAQULTRMMCSEARIDMNRVRLGQLTDRDFSRSLVDVASRLSEAP-IYIDTDPDLTL 295

Query: 266 GKNLIPAILDSMISKYRPSVVGIDQLSLMS--ESYPSREQKRIQYANITMDLYKISAKYG 323
 + A ++S+ + ++ ID L LMS S S E ++ + A I+ L ++ + G
 Sbjct: 296 ME--VRARARLLVSQNVGLIITDYLQMLSGPGSGKSGENRQGEIAAISRLKALARELG 353

Query: 324 IPIVLNVQAGRSKTEGAESMELEHIAESDGVQGNASRVIAMKRD-----EKSGILEL 376
 IPI+ Q R+ + + L + ES + Q+A V+ + RD EK+GI E+
 Sbjct: 354 IPIIALSQLGRAVEARPNKRPMLSDLRESGSIEQDADLVMIYRDEYYPHSEKAGIAEI 413

Query: 377 SVVKHRYG 384
 V K R G
 Sbjct: 414 IVGKQRNG 421

Query= sid|114831|lan|dplORF010 Phage dpl ORF|8699-9859|2
 (386 letters)

>gi|2760912 (AF037258) RecA protein [Chlorobium tepidum]
 Length = 346

Score = 133 bits (331), Expect = 2e-30
 Identities = 99/340 (29%), Positives = 164/340 (48%), Gaps = 66/340 (19%)

Query: 44 GGLPRKRVVEFFPGPSSGKTTLSALDIIVNAQMVFXXXXXXXXXXXXXXXXXXNARASKASKY 103
 GGLPR RV E +GPSSGKTT AL + AQ
 Sbjct: 67 GGLPRGRVTEIYGPSSGKTTLALHAIAEAQ-----XNG 100

Query: 104 AVKLEMLQDLSLQEPKIVYLDLENTLDTEWAKKIGVDVDNIWIVRPEMNSAEBILQYVL 163
 + L +D E+ D +A+K+GVD++ + + +PE S E+ L V
 Sbjct: 101 GIAAL-----VDAEHAFDPTYARKLGVGINALLVSQPE--SGEQALSIVE 143

Query: 164 DIFETGEVGLVVLOSLPVMVSNLIDEELTKAYAGISAPLTFESRKVTPLLLTRYNAIFL 223
 + +G V ++V+DS+ +V Q ++ E+ + +++ RK+T +++ ++ L
 Sbjct: 144 TLVRSGAVDIIVIDSVAALVPOAELEGEMGDSVVGQLQARLMSQALRKLITGAISKSSSVCL 203

Query: 224 GINQIREDMNSQYNA-YSTPGKMMWKAACAVRLKFRKGDYLDENGASLTTRTARNPAGNVV 282
 INQ+R+ + Y + +T GGK K +VRL RK + ++G L GN
 Sbjct: 204 FINQLRDKIGVMYGSPETTTGGKALKFYSSVRLDIRKIAQI-KDGEELV-----GNRT 255

Query: 283 SSFVEKTKAPKPDRLKLVSYTLSDGIQIENDLVDAVEFGVIOKAGAWFSIVDLETGEI 342
 + V K K P R + + Y +GI + +L+D+AVEFG+I+K+GAWFS + G
 Sbjct: 256 KVKVVKVKV-APPFKTAEFDILYGEISVLGELIDLAVEFGIINKSGAWFSYGTETKLG-- 312

Query: 343 MTEDESEPLKFGQKANLVRRFKEDDYLFDMVMVAVHEIIT 382
 QG+ N+ + KED+ L + + V +++T
 Sbjct: 313 -----QGRENVKLLKEDETIRNTIRQQVRDMLT 341

Query= sid|114832|lan|dplORF011 Phage dpl ORF|28017-29096|3
 (359 letters)

>gi|2444110 (U88974) ORF31 [Streptococcus thermophilus temperate bacteriophage
 01205]
 Length = 348

Score = 187 bits (469), Expect = 1e-46
 Identities = 118/358 (32%), Positives = 187/358 (51%), Gaps = 21/358 (5%)

Query: 3 IYDYNAGSIASYIQALPSNALQYLGPTLFPAQQTGDISWLKGANNLPVTIOPSNYDA 62
 IYD + A IA Y AL N LG ++FP +Q GT +S++KGA+ V ++ + +D
 Sbjct: 4 IYDKVTASNIAGYFNALQENVSSLTGESIFPARKQLGTLKLSYIKGASGQSVALKAAAFDT 63

Query: 63 KASLRERAGFSKQATEMAFFRESMRIGEDKDRNLQMLLNQSSA-LAQPLITQLYNDTKNL 121

404

```

      ++R+R      +M FF+E+M + E DRQ L ++ + +A L ++ ++ND L
Sbjct: 64  NVTIRDRVSAEMHDEQMPFFKEAMLVKENDRQQLNLVKDSGNAVLVNTIVAGIFNDNLTL 123

Query: 122  VDGVEAQAEYMRMQLLQYKFTVKSTNSEAQYTYDYNMDAKQYAVTKKWTNPASSOPIA 181
      V+G A+ E MRMQ+L GK S Y D K+Q V+K W P + P+A
Sbjct: 124  VNGARARLEAMRMQVLATGKIAFTSDGVNKKDIDYGVKPDHKKQ--VSKSWAEPG-ATPLA 180

Query: 182  DILAAMDDIENRTGVRPRTMVLNRNTYNQMTKSDSIKKAL-AIGVQGSWENFLLASDAE 240
      D+ A+ + G+ P R V+N T+ + K+ S K + + GS + ++ E
Sbjct: 181  DLEDAI-ETARELGLNPERAVMNAKTFGLIRKAASTVKVIKPLAGDGS----AVTKAELE 235

Query: 241  KFIAEKTGLQIAVYSKKIAQFADADKLPDVGNIQFNLIIDDGKVVLLPPDAVGHTWYGT 300
      +IA+ G+ I + + D G + +F DG + L+P +G+T +GTT
Sbjct: 236  NYIADNFGVSIVLENGTYRN-----DKGEVSKF--YPDGHLLTLPNGPLGNTVFGTT 285

Query: 301  PEAFLDASGGT-DAQVQVLGGFTVTITYLEKHPVNIATVVSAMVIPSFEGIDYVGVLT 357
      PE DL + T +A+V+++ G VTT PVN+ T VS V +PSFE +D V +LT
Sbjct: 286  PEESDLFADNTVNAEVEIVDNGIAVTTTCTDTPVNVQTKVSMVALPSPERLDDVYMLT 343

```

Query= sid|114834|lan|dp1ORF013 Phage dp1 ORF|10215-11240|3
(341 letters)

>sp|P09122|DP3X_BACSU DNA POLYMERASE III SUBUNITS GAMMA AND TAU
Length = 563

Score = 182 bits (458), Expect = 2e-45
Identities = 118/353 (33%), Positives = 176/353 (49%), Gaps = 31/353 (8%)

```

Query: 7   YRPQTFEEVVAQEYVKEILLNQLQNGAIKHGVLFCXXXXXXXXXXXXXIFAKDVN----- 60
      +RPQ FE+VV QE++ + L N L H YLF +IFAK VN
Sbjct: 10  FRPQRFEDVVGQEHITKTILQNALLQKCFSHAYLFSGPRGTGKTSAAKIFAKAVNCEHAPV 69

Query: 61  -----KGL-----GSPIEIDAASNNGVENVNIIEDSRYKSMDSFVKVYIIDEVH 105
      KG+ IEIDAASNNGV+ +R+I + ++ +KVYIIDEVH
Sbjct: 70  DEPCNECAACKGITNGSISDVIEIDAASNNGVDEIRDIDKVKFAPSAVTYKVYIIDEVH 129

Query: 106  MLSTGAFNALLKLEEPSGGTVFILCTDPQKIPDTILSRVQRFDPTRIDNDIVNQLQF 165
      MLS GAPNALLKLEEP +FIL TT+P KIP TI+SR QRDF RI + IV ++
Sbjct: 130  MLSIGAFNALLKLEEPPEHCIFILATTEPHKIPLTIISRCQRDFDKRITSQAIVGRMNK 189

Query: 166  IIESENEEGAGVSYERDALSFIGKLANGGMRDSITRLEKVLVDYSHHVDMEAVSNAL---G 222
      I+++E E +L I A+GGMRD+++ L+++ +S D+ V +AL G
Sbjct: 190  IVDAEQ-----LQVEEGSLEIIASAAHGGMRDALSLLDQAISFSG--DILKVEDALLITG 242

Query: 223  VPDYETFASLVEAIAANYDGSKCLEIVNDFHYSKDKLKLVTNFTDFLLEVCKYNLVRDIS 282
      L +++ + + S LE +N+ GKD + + + ++ Y +
Sbjct: 243  AVSQLYIGKLAKSLHDKNVSDALETNLNLLQQKDPAKLIEDMIFYFRDMLLYKTAPGLE 302

Query: 283  ITQLPAHFESKLEQFCEAFQYPTLLWMLLEEMNELAGVVKWSPNAKPIIETKLL 335
      + + E L M++ +N+ +KM + + E ++
Sbjct: 303  GVLEKVKVDETFRELSQIPQAQALYEMIDILNKSHQEMKWTNHPRIFFEVAVV 355

```

Query= sid|114835|lan|dp1ORF014 Phage dp1 ORF|50961-51974|3
(337 letters)

>sp|P47492|PRIM_MYCGE DNA PRIMASE >gi|1361496|pir|F64227 DNA primase (dnaE) homolog
MG250 - Mycoplasma genitalium (SGC3) >gi|3844848
(U39704) DNA primase (dnaE) [Mycoplasma genitalium]
Length = 607

Score = 57.0 bits (135), Expect = 2e-07
Identities = 53/190 (27%), Positives = 89/190 (45%), Gaps = 17/190 (8%)

```

Query: 146  EELDKYRFIHP-----YMYERKLTDELIEMFDVGYDK--LHDCITFPVRNLKGETVFF 196
      E +++Y FI+P Y++ K + + FD K + I P+ + G V F
Sbjct: 170  ESMERYPFPIPKIKPSELYLFS-KTNQQGLGFFDFNTKKATFQNGQIMIPIDFNGNPVGF 228

Query: 197  NRRSVRSKPHQYGEDDPKTEFLYGQYELVAFRDYFEKPISQVFVTESVINCLTWSMKIP 256
      + RSV + ++ EF + + EL+ K ++Q+F+ E + TL + K
Sbjct: 229  SARSDVNINKLKYKNSADHEF-FKKGELLFNHRLNQLNQLFIVEGYFDVFTLTNSKFE 287

Query: 257  AVALMGVGGGN-QINLLKR--LPYRNIVLALDPDNAGQTAQSKLYRQLKRSK-VVRFLNY 312
      AVALMG+ + QI +K + +VLALD D +GQ A L +L + +V + +
Sbjct: 288  AVALMGLALNDVQIKAIKAHFKEQLTLVLALDNDASGQNAVPSLIEKLNNNNFIVEIVQW 347

```

405

Query: 313 PKEFYDNKWD 322
 + D WD
 Sbjct: 348 EKNYKD--WD 355

Query= sid|114837|lan|dp1ORF016 Phage dp1 ORF|43413-44303|3
 (296 letters)

>emb|CAB07986| (Z93946) N-acetylmuramoyl-L-alanine amidase [bacteriophage Dp-1]
 Length = 296

Score = 661 bits (1686), Expect = 0.0
 Identities = 296/296 (100%), Positives = 296/296 (100%)

Query: 1 MGVDIEKGVAMQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGASSAGWAVNTEYMH 60
 MGVDIEKGVAMQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGASSAGWAVNTEYMH
 Sbjct: 1 MGVDIEKGVAMQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGASSAGWAVNTEYMH 60

Query: 61 AWLIENGYELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIIHCNYAYDGIS 120
 AWLIENGYELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIIHCNYAYDGIS
 Sbjct: 61 AWLIENGYELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIIHCNYAYDGIS 120

Query: 121 VNDHDERWYYAGQPYYYVYRLTNANAQPAEKKLGWQKDATGFYARANGTYPKDEFEYIE 180
 VNDHDERWYYAGQPYYYVYRLTNANAQPAEKKLGWQKDATGFYARANGTYPKDEFEYIE
 Sbjct: 121 VNDHDERWYYAGQPYYYVYRLTNANAQPAEKKLGWQKDATGFYARANGTYPKDEFEYIE 180

Query: 181 ENKSWFYFDDQGYMLAEKWLKHTDGNWYWFDRDGYMATSNKRIGESWYYPNRDGSMTGW 240
 ENKSWFYFDDQGYMLAEKWLKHTDGNWYWFDRDGYMATSNKRIGESWYYPNRDGSMTGW
 Sbjct: 181 ENKSWFYFDDQGYMLAEKWLKHTDGNWYWFDRDGYMATSNKRIGESWYYPNRDGSMTGW 240

Query: 241 IKYYDNWYYCDATNGDMKSNAPIRYNDGWYLLLPDGLADKPQFTVEPDGLITAKV 296
 IKYYDNWYYCDATNGDMKSNAPIRYNDGWYLLLPDGLADKPQFTVEPDGLITAKV
 Sbjct: 241 IKYYDNWYYCDATNGDMKSNAPIRYNDGWYLLLPDGLADKPQFTVEPDGLITAKV 296

Query= sid|114841|lan|dp1ORF020 Phage dp1 ORF|1864-2658|1
 (264 letters)

>emb|CAB13247| (Z99111) similar to coenzyme PQQ synthesis [Bacillus subtilis]
 Length = 243

Score = 217 bits (548), Expect = 5e-56
 Identities = 117/248 (47%), Positives = 163/248 (65%), Gaps = 15/248 (6%)

Query: 23 MPIMEIFGPTIQEGMVGQKTIFIRTGGCDYHCNWCDSAPFTWNGTTEPE--YITGKEAA 80
 +P++EIFGPTIQEGMVGQKT+F+RT GCDY C+WCDSAFTW+G+ + + ++T +E
 Sbjct: 5 IPVLEIFGPTIQEGMVGQKTIMFVRTAGCDYSCSWCDSAFTWDGSAKKDIRMWTAEIF 64

Query: 81 SRILKLAFNKGEQICNVHTLTGGNPALINEPMKMSILKEHGPKFGLTQGTTRFQENF 140
 + + D G +HVT++GGNPAL+ + + I +LKE+ + LETQGT +Q+WP
 Sbjct: 65 ASL-----KDIGGDAPSHVTISGGNPALLKQ-LDAFIELLKENNIRAALETQGTQVYQDWF 118

Query: 141 KEVSDITISPKPPSSGMRTNMKILEAIVDRM--NDENLDWSPKIVIPDSNDLAYARDMPK 198
 + D+TISPKPPSS M TN + L+ I+ + ND S K+VIF++ DL +A+ + K
 Sbjct: 119 TLIDDLTISPKPPSSKMVTNPFQKLDHILTSIQENDRQHAVSLKVIFNDEDLFAKTVHK 178

Query: 199 TFEGLRFPVNYLSVGNANAY--EEGKISDRLLLEKLGWLNDKVYEDPAFNNVRPLPQLHTL 256
 + G YL VGN + + ++ + LL K L DKV D N VR LPQLHTL
 Sbjct: 179 RYPG---IPFYLVQGNDDVHTTDDQSLIAHLLGKYEALVDKVAVDABLNLVRVLPQLHTL 235

Query: 257 VYDNKRGV 264
 ++ NKRGV
 Sbjct: 236 LWGNKRGV 243

406

Query= sid|114842|lan|dp1ORF021 Phage dp1 ORF|2504-3295|2
(263 letters)

>sp|P19465|GCH1_BACSU GTP CYCLOHYDROLASE I (GTP-CH-I) >gi|98411|pir||A38256 GTP
cyclohydrolase I (EC 3.5.4.16) - Bacillus subtilis
>gi|143231 (M37320) regulatory protein [Bacillus
subtilis] >gi|143799 (M80245) MtrA [Bacillus subtilis]
>gi|2634696|emb|CAB14194| (Z99115) GTP cyclohydrolase I
[Bacillus subtilis]
Length = 190

Score = 208 bits (523), Expect = 4e-53
Identities = 103/185 (55%), Positives = 133/185 (71%), Gaps = 1/185 (0%)

Query: 80 VTLDNTEAAVQRLFLGLGEDAERDGLQDTFFRFVKALAEHTVGYPREDPKLHLEKTFDVDH 139
V + E AV+++ +GED R+GL DTP R K AE G EDPK H + F +H
Sbjct: 4 VNKEQIEQAVRQILEAIGEDFNREGLLDTPKRVAKMYAEVFSGLNEDPKENHQTIFGENH 63

Query: 140 EDLVLVKDI PFNSLCEHHLAPFVGKVHIAIYIPKD-KITGLSKFGRVVEGYAKRLQVQERL 198
E+LVLVKDI F+S+CEHHL PF GK H+AYIP+ K+TGLSK R VE AKR Q+QER+
Sbjct: 64 EELVLVKDIAFHSMCEHHLVPFYGKAHVAYIPRGKVTGLSKLARAVEAVAKRPQLQERI 123

Query: 199 TQOIADAIQEVLPQAVAVIVEAEHTCMISGRGIKKHGAITVTSTMRGLFQDDASARAELL 258
T IA++I E L+P V V+VEAEH CM+ RG++K GA TVTS +RG+F+DDA+ARAE+L
Sbjct: 124 TSTIAESIVETLDPHGVMVVVEAEHMCMTMRGVRKPGAKTTSVAVRGVFKDDAAARAELV 183

Query: 259 QLIXK 263
+ IK+
Sbjct: 184 ZHIKR 188

Query= sid|114843|lan|dp1ORF022 Phage dp1 ORF|30896-31675|2
(259 letters)

>gi|2347102 (U77367) internalin [Listeria monocytogenes]
Length = 821

Score = 55.0 bits (130), Expect = 5e-07
Identities = 44/149 (29%), Positives = 63/149 (41%), Gaps = 13/149 (8%)

Query: 119 FRMNIYVPNYVG--DSIVNVVKITLNNCTGKAPGLSIGKEFYAPEFNIKAREATKAGLPV 176
F + VPN + D + + NN T AP L Y PE +K + K +
Sbjct: 383 FSKTSLSVPMNITSIDGTLIAPETISNNGTYDAPNLKWSLPNYLPE--VKYTFSSQKIPGT 440

Query: 177 KSMQVYAQLPAVLV-----RVTFDLNGGTGTADAVRVEAGKKISPKPVDPTLTGKAPKGW 231
+ +Y + L+ +VTF++ G T + V E + P+P PT G F GW
Sbjct: 441 GTSNYSGPITQPLKELLDYKVTFFNVEGNTSEVETVTEE---NLIPEPTSPKQGYTFDQW 497

Query: 232 -KVEGESTIWDFFDNHMPDRDVKLVAQPA 259
E T WDF MP D+ L A F+
Sbjct: 498 YDAETGGTKWDFTTGQMPANDLTLIAHFS 526

Score = 43.4 bits (100), Expect = 0.002
Identities = 47/195 (24%), Positives = 73/195 (37%), Gaps = 12/195 (6%)

Query: 72 YDLTFKDNFTDPEIMALIEGGTVRQQGGTIAGYDT-PMLAQGASNMKPFPMNIYVPNY-- 128
YD + T + +G + GG + T M A + F +N Y N+
Sbjct: 547 YDALLNEPTTPTKQGYTFDQWYDAETGGNKWDFKTKMPANDVAFYAHFTINNYQANFDI 606

Query: 129 ---VGDSIVNVVKITLNNCTGKAPGLSIGKEFYAPEFNIKAREATKAGLPVKSMDYVAQL 185
V + + Y + T G + + A K TK +P + A
Sbjct: 607 DGEVKNETIAYDTLLNEPTTPTKQGYTFDQWYDAETGGTKWDFKKE-MPANDVTLIAHF 665

Query: 186 PAVLRRVTFDLNGGTGTADAVRVEAGKKISPKPVDPTLTGKAPKGW-KVEGESTIWDFFDN 244
+ FD++G T + V +A + P+P P+ TG +GW E T WDF
Sbjct: 666 TINNYQANFDIDGAV-TREVVNYDA---LIPEPTSPSKTGFTLEGWYDAEVGGTKWDFKT 721

Query: 245 HMMPPDRDVKLVAQPA 259
MP D+ L A F+
Sbjct: 722 MKMPANDITLYAHFS 736

Score = 38.3 bits (87), Expect = 0.057
Identities = 42/169 (24%), Positives = 59/169 (34%), Gaps = 10/169 (5%)

407

Query: 96 QGGTIAGYDT-PMLAQGASNMKPFMMNIYVPNYVGDIVNYVKIT----LNNCTGKAPG 150
 + GGT + T M A + F +N Y N+ D +V + LN T
 Sbjct: 501 ETGGTKWDFTTGQMPANDLTLYAHFVSNSYQANFDIDGVVTNEAVVYDALLNBPPTPTKQ 560

Query: 151 LSIGKEFYAPEFNKAREATKAGLPVKSMQYVAQLPAVLRRVTFDLNGGTCTADAVRVEA 210
 +Y E + +P + + A + FD++G A
 Sbjct: 561 GYTFDGYDAETGGNKWDFKTMKMPANDVAFYAHFTINNYQANFDIDGEVQNETI----A 616

Query: 211 GKKISPKPVDPTLTGKAFKGW-KVEGESTIWDNFMMPDRDVKLVAQF 258
 + +P PT G F GW E T WDF MP DV L A F
 Sbjct: 617 YDTLLNEPTTPTKQGYTFDGYDAETGGTKWDFKTKEMPANDVTLYAHP 565

Query= sid|114850|lan|dp1ORF029 Phage dpl ORF|662-1348|2
 (228 letters)

>gi|2650185 (AE001074) succinoglycan biosynthesis regulator (exsB)
 [Archaeoglobus fulgidus]
 Length = 239

Score = 119 bits (295), Expect = 2e-26
 Identities = 79/224 (35%), Positives = 113/224 (50%), Gaps = 11/224 (4%)

Query: 1 MKSVLLSGGVDSATCLAEVDKMGSKNVHAIAPNYGQKHEAELENAANVAMFYGVKPTI 60
 MK+V+LLSGG+DS+T L +D G VHA+ F YGQKH E+E+A VA V+
 Sbjct: 1 MKAVMLLSGGIDSSTLLYLLD--GGYEVHALTFPYGQKHSKEIESAEKVAKVVRHLK 58

Query: 61 LEIDSKIYXXXXXXLLQKGEISHGKSYAEILAEKEVVDVTPFRNGLMLSQXXXXXXX 120
 ++I S I+ L G+ E+ Y+E + + T VP RN ++LS
 Sbjct: 59 VDI-STIHDLISYGALTGESEVPKA-FYSSEVQR-----TIVPNRMILLS--IAAGYAV 110

Query: 121 XXXXXXXXXXXXXXXXXXXXPDCTPEFYNSMSNAMEYGT-GGKVTLVAPLLTLTKAQVVW 179
 PDC EF ++ A+ V + AP + +TKA +V+
 Sbjct: 111 KIGAKEVHYAAHLSDYSIYPCDKKEFVKALDTAVYLANIWTPEVRAPFVDMTKADIVRL 170

Query: 180 GIDLDVPYFLTRSCYSDAESCGTCATCIDRKKAFENGMTDPI 223
 G+ L VPY LT SCYE C +C TC++R +AF NG+ DP+
 Sbjct: 171 GLKLGVPYELTWSCYEGGDRPCLSCGTCLETRAPLANGVKDPL 214

Query= sid|114855|lan|dp1ORF034 Phage dpl ORF|131-652|2
 (173 letters)

>emb|CAB13248| (Z99111) similar to hypothetical proteins [Bacillus subtilis]
 Length = 165

Score = 220 bits (556), Expect = 4e-57
 Identities = 103/139 (74%), Positives = 117/139 (84%)

Query: 5 TTRTDABLTVGVTLLGNQDTKYDYNDYNDVLETFPNKHPPENNYLVTFDGYEFTSLCPKTQ 64
 TTR ++EL GVTLGNQ T Y ++Y PDVLE+FPNKH +Y V F+ EFTSLCPKTQ
 Sbjct: 2 TTRKESLEGVTLGNQGTNYLFEYAPDVLESFPNKHVNRDYFVKPNCPEFTSLCPKTQ 61

Query: 65 PDFANVFISYIPNEKQVESKSLKLYLFSFRNHGDFHEDCMNIIINDLYELMEPKYIEVMG 124
 PDFA ++ISYIP+EKQVESKSLKLYLFSFRNHGDFHEDCMNII+NDL ELM+P+YIEV G
 Sbjct: 62 PDFATIIYISYIPDEKQVESKSLKLYLFSFRNHGDFHEDCMNIIIMNDLIELMDPRYIEVMG 121

Query: 125 LFTPRGGISIIYFPVNVNP 143
 FTTPRGGISI P+ N P
 Sbjct: 122 KFTPRGGISIDPYTYNGKP 140

Query= sid|114857|lan|dp1ORF036 Phage dpl ORF|48808-49362|1
 (184 letters)

>gi|1353529 (U38906) ORF12 [Bacteriophage rlt]
 Length = 296

Score = 53.5 bits (126), Expect = 1e-06
 Identities = 42/149 (28%), Positives = 70/149 (46%), Gaps = 9/149 (6%)

Query: 34 IASNTVNGKTSWAVRLLQRYLAETALDGRIVEKGMFVVSQALLTEFGDYNYFQTMQEFL 93
 + S G GK+ A+ +L+ L T L ++ V + P + + F + + P+
 Sbjct: 155 VVSGPAGTGKSHLAMSILKQCLQHTDLT--VIFASWSEVLHLINKDSFONKDSFYSTRYFM 212

408

Query: 94 ERFERLKTCELLVIDEIGGGSLTKASYFYLYDLVNYRVNDNLSTIYTTNYTDDDEIIDLLG 153
 E F + +LLVID+IG +T+ S L +++ R TI TIN DEI
 Sbjct: 213 EVF---RNTDLLVIDDIGSEKITEWSMSLLTEVLDART----KTIITNLKSDEIRKKYH 265

Query: 154 QRLYSRIYDTSVVLDQFQASNVRLGVSEI 182
 R YSR++ F N++ VS++
 Sbjct: 266 NRTYSRLFRGIGKAPNPFENIKDKRVSQL 294

Query= sid|114859|lan|dp1ORF038 Phage dp1 ORF|1350-1871|3
 (173 letters)

>sp|P44123|YB90_HAEIN HYPOTHETICAL PROTEIN HI1190 >gi|1074675|pir||F64021 hypothetical
 protein HI1190 - Haemophilus influenzae (strain Rd KW20)
 >gi|1574117 (U32798) 6-pyruvoyl tetrahydrobiopterin
 synthase, putative [Haemophilus influenzae Rd]
 Length = 141

Score = 100 bits (247), Expect = 6e-21
 Identities = 59/143 (41%), Positives = 83/143 (57%), Gaps = 10/143 (6%)

Query: 2 RVSKTLTFDAAHQLVGHFGKCANLHGHTYKVEISLAGGTYDHGSSQGMVVDFFYHVKKIA- 60
 ++SK +FD AH L GH GKC NLHGHTYK+++ ++G Y G+ + MV+DF +K I
 Sbjct: 3 KISKEFSFDMAHLLDGHGDKCQNLHGHTYKLVQVEISGDLTKSGAKAMVIDPSDLKSIVK 62

Query: 61 GTFIDRLDHAVLL-QGNP-----IALNAVDTKRVLPFGFRITAEENMSRPLTWTLTELMWK 115
 +D +DHA + Q NE L +++K FRITAE ++RF+ L +
 Sbjct: 63 KVILDPMDHAFIYDQTNERESQIATLLQKLNSTKFGVPFRITAEIARFIFNRLKH--DE 120

Query: 116 HARIDSIKLWETPTGCAECTYYE 138
 I SI+LWETPT + C Y E
 Sbjct: 121 QLSISSIRLWETPT--SFCEYQE 141

Query= sid|114860|lan|dp1ORF039 Phage dp1 ORF|3306-3803|3
 (165 letters)

>emb|CAA68244| (X99978) ORF7: hydrophobic protein [Lactobacillus plantarum]
 Length = 168

Score = 64.4 bits (154), Expect = 5e-10
 Identities = 49/156 (31%), Positives = 84/156 (53%), Gaps = 9/156 (5%)

Query: 8 MLVRTALIAALYVTLTVAFSAISY--GPIQFRVSEALILLPLNHRKWTGIVLGTIIANF 65
 W++ AL+AA+YV L + +A S G IQFRVSE L L ++N ++ GIV G I+ +
 Sbjct: 9 WIIN-ALVAMMYVVLCLGPAAFSLASGAIQFRVSEGLNHLAVFNKRYKNGIVAGVILFDA 67

Query: 66 FSP-LGLIDVLPGLSLATPLGXXXXXXXXXXSPLYSLICPVLA---NAYLIALELRIVY 120
 F P L++VLF + L ++ + +A + ++IAL + ++
 Sbjct: 68 FPGASLLNVLPFGGQSLLALLVLTWLPKLTWVQMLNLIALFTVSMFIALMITMS 127

Query: 121 S-LPPWESVIYVGISEAIIVLISYFLISTLAKNHF 155
 S + FW + + +SE II+ I+ ++ +L + HF
 Sbjct: 128 SGVAFNPYTLTALSELIIMSITAFIMYSLDRVLHF 163

Query= sid|114862|lan|dp1ORF041 Phage dp1 ORF|8208-8699|3
 (163 letters)

>gi|2522313 (AF012906) dUTPase homolog [Bacillus subtilis]
 >gi|2634394|emb|CAB13893| (Z99114) similar to
 deoxyuridine 5'-triphosphate nucleotidohydrolase
 [Bacillus subtilis] >gi|3025643 (AF020713) putative
 dUTPase [Bacteriophage SPBc2]
 Length = 142

Score = 108 bits (267), Expect = 2e-23
 Identities = 65/160 (40%), Positives = 83/160 (51%), Gaps = 25/160 (15%)

Query: 5 VDVKMIDPKLDRLKYT--GDWVDVRISITKIDADSADVSRCRKVLQKAQVYSVAAGECI 62
 + +K +D R+ GDN+D+R + I D +
 Sbjct: 3 IKIKYLDDETQTRINKMEQGDWIDLRAEDVAIKKDEFKL----- 41

Query: 63 KIAHGPALELPKGYEAILHPRSSLFKKTGILFVSS-GVIDEGYKGTDEWFSVMYATROA 121
 + G A+ELP+GYEA + PRSS +K G+I +S GVIDE YKGD D WF YA RD
 Sbjct: 42 -VPLGVAMELPEGYEAHVVPSSSTYKNFGVIQTNSMGVIDESYKGDNDFWFFPAYALRDT 100

Query: 122 DIFYDQRIAQFRIQEKPAPKFNFVESLGNARGGHGSTG 161
I RI QFRI +K PA+ V+ LGN RGGHGSTG
Sbjct: 101 KIKKGDRCIFRIMKKMPAVDLIEVDRLGNGDRGGHGSTG 140

Score = 287 bits (728), Expect = 2e-77
Identities = 142/142 (100%), Positives = 142/142 (100%)

Query: 121 VEALYEKYKKLP IREEDLDETI 142
VEALYEKYKKLP IREEDLDETI
Sbjct: 121 VEALYEKYKKLP IREEDLDETI 142

Score = 147 bits (367), Expect = 1e-35
Identities = 75/75 (100%), Positives = 75/75 (100%)

Query: 61 EQKLRETRYAIED EI 75
EQKLRETRYAIED EI
Sbjct: 61 EQKLRETRYAIED EI 75

Score = 63.2 bits (151), Expect = .2e-10
Identities = 34/74 (45%), Positives = 34/74 (45%)

Query: 61 NYQKEQEAQNNEVE 74
NYQKEQEAQNNEVE
Sbjct: 61 NYQKEQEAQNNEVE 74

Condensed listing of homology information from above

Phage: dpl

Database: nr

Program: Blastp

Query= sid|114822|lan|dplORF001 Phage dpl ORF|36698-40390|2
(1230 letters)

gi 2444124	(U88974) ORF45 [Streptococcus thermophilus temperate ...	467	e-130
gi 928828	(L44593) ORF1904; putative [Lactococcus lactis phage B...	427	e-118
gi 2935676	(AF032121) unknown [Streptococcus thermophilus bacter...	309	1e-82
gi 2935691	(AF032122) unknown [Streptococcus thermophilus bacter...	306	7e-82
gi 3540289	(AF057033) putative anti-receptor [Streptococcus ther...	279	6e-74
gi 4530154	[gb AAD21894.1 (AF085222) putative tail-host specific...	220	3e-56
gi 930045	[emb CAA33387 (X15332) alpha-1 (III) collagen [Homo sa...	58	4e-07
gi 1070603	[pir CGHU7L collagen alpha 1(III) chain precursor - h...	58	4e-07
gi 4502951	[ref NP_000081.1 PCOL3A1 collagen, type III, alpha 1 ...	58	4e-07
gi 115290	[sp P04258 CA13_BOVIN COLLAGEN ALPHA 1(III) CHAIN >gi 7...	58	4e-07
gi 575322	[emb CAA36279 (X52046) type III collagen [Mus musculus]	57	8e-07
gi 2119163	[pir S59856 collagen alpha 1(III) chain precursor - m...	57	8e-07
gi 543912	[sp P13941 CA13_RAT COLLAGEN ALPHA 1(III) CHAIN >gi 543...	57	1e-06
gi 3171998	[emb CAA06510 (AJ005395) collagen alpha 1 (III) [Ratt...	57	1e-06
gi 3947565	[emb CAA90250 (Z49967) similar to collagen; cDNA EST ...	54	7e-06
gi 423403	[pir A46053 bullous pemphigoid antigen, BPAG2, type XV...	53	9e-06
gi 115410	[sp P12114 CCS1_CAEEL CUTICLE COLLAGEN SQT-1 >gi 84437 ...	53	9e-06
gi 3873801	[emb CAA90084 (Z49907) cuticle collagen SQT-1; cDNA E...	53	9e-06

Query= sid|114823|lan|dplORF002 Phage dpl ORF|32386-35835|1
(1149 letters)

gi 3341922	[dbj BAA31888 (AB009866) orf 15 [bacteriophage phi FVL]	280	3e-74
gi 4126622	[dbj BAA36642.1 (AB016282) ORF36 [bacteriophage phi-105]	232	1e-59
gi 1369948	[emb CAA59194 (X84706) host interacting protein [Bact...	201	3e-50
gi 3139112	(AF063097) gpt [Bacteriophage P2]	188	2e-46
gi 3337272	(U32222) G protein [Bacteriophage 186]	161	3e-38
gi 4063799	[dbj BAA36253 (AB008550) orf25; similar to T gene of ...	159	8e-38
gi 3172274	(AF022214) minor tail subunit; putative tape-measure ...	123	6e-27
gi 465127	[sp Q05233 VG26_BPM15 MINOR TAIL PROTEIN GP26 >gi 41904...	108	2e-22
gi 3540284	(AF057033) putative minor tail protein [Streptococcus...	99	2e-19
gi 2444119	(U88974) ORF40 [Streptococcus thermophilus temperate ...	90	6e-17
gi 2634555	[emb CAB14053 (Z99115) yomI [Bacillus subtilis] >gi 3...	66	1e-09
gi 2392838	(AF011378) unknown [Bacteriophage sk1]	64	5e-09
gi 2764873	[emb CAA66557 (X97918) gene 18.1 [Bacteriophage SPPI]	62	3e-08
gi 1353559	(U38906) ORF42 [Bacteriophage rlt]	61	6e-08
gi 630841	[pir S39079 puff C-8 protein - fungus gnat (Rhynchosci...	55	2e-06
gi 1730865	[sp P51731 YO27_BPHF1 HYPOTHETICAL 72.8 KD PROTEIN IN ...	53	8e-06
gi 224288	[prf 1101273J ORF 7 [Bacteriophage HP1]	53	1e-05

Query= sid|114824|lan|dplORF003 Phage dpl ORF|53538-55877|3
(779 letters)

gi 118825	[sp P00582 DPO1_ECOLI DNA POLYMERASE I (POL I) >gi 6705...	193	3e-48
gi 2982102	[pdb 1KPS A Chain A, All-Oxygen Dna Complexed To The 3...	193	3e-48
gi 229889	[pdb 1DPI DNA Polymerase I (Klenow Fragment) (E.C.2....	193	3e-48
gi 1169402	[sp P43741 DPO1_HAEIN DNA POLYMERASE I (POL I) >gi 107...	191	1e-47
gi 2688462	(AE001156) DNA polymerase I (polA) [Borrelia burgdorf...	190	3e-47
gi 809180	[pdb 1KLN A Escherichia coli	190	3e-47
gi 1913934	[emb CAA72997 (Y12328) DNA-directed DNA polymerase I ...	189	8e-47
gi 4090935	(AF028719) DNA polymerase type I [Rhodothermus sp. 'I...	175	1e-42
gi 4731571	[gb AAD28505.1 AF121780_1 (AF121780) DNA polymerase I ...	174	2e-42
gi 1633576	(U57757) similar to proofreading 3'-5' exonuclease an...	173	4e-42
gi 3322368	(AE001195) DNA polymerase I (polA) [Treponema pallidum]	172	9e-42
gi 1006595	[dbj BAA10748 (D64005) DNA polymerase I [Synechocysti...	171	2e-41
gi 585062	[sp Q07700 DPO1_MYCTU DNA POLYMERASE I (POL I) >gi 4161...	163	5e-39
gi 4376908	[gb AAD18751 (AE001645) DNA Polymerase I [Chlamydia p...	157	2e-37
gi 1169403	[sp P46835 DPO1_MYCLE DNA POLYMERASE I (POL I) >gi 107...	152	7e-36
gi 2145839	[pir S72949 DNA polymerase I - Mycobacterium leprae >...	152	7e-36
gi 1405438	[emb CAA67184 (X98575) DNA-dependent DNA polymerase (...	152	9e-36
gi 2506365	[sp P80194 DPO1_THECA DNA POLYMERASE I, THERMOSTABLE (...	147	2e-34
gi 3328929	(AE001322) DNA Polymerase I [Chlamydia trachomatis]	147	3e-34

411

gi 3913510 sp O52225 DPO1_THEFI DNA POLYMERASE I, THERMOSTABLE (...)	146	7e-34
gi 1205984 (U33536) DNA polymerase I [Bacillus stearothermophilus]	146	7e-34
gi 118827 sp P13252 DPO1_STRPN DNA POLYMERASE I (POL I) >gi 9802...	145	9e-34
gi 1942202 pdb 1JXE Stoffel Fragment Of Taq Dna Polymerase I	145	1e-33
gi 1943520 pdb 1KTQ Dna Polymerase	145	1e-33
gi 1084022 pir JX0359 DNA-directed DNA polymerase (EC 2.7.7.7) ...	145	1e-33
gi 507891 dbj BAA06775 (D32013) DNA Polymerase (Thermus aquaticus)	145	1e-33
gi 118828 sp P19821 DPO1_THEAQ DNA POLYMERASE I, THERMOSTABLE (T...	145	1e-33
gi 1706502 sp P52028 DPO1_THETH DNA POLYMERASE I, THERMOSTABLE (...)	144	2e-33
gi 1097211 prf 2113329A DNA polymerase [Thermus aquaticus therm...	144	2e-33
gi 2098289 pdb 1TAU A Chain A, Structure Of Dna Polymerase	143	3e-33

Query= sid|114825|lan|dp1ORF004 Phage dpl ORF|40401-42440|3
(679 letters)

gi 1934761 emb CAB07981 (Z93946) hypothetical protein [bacterio...	1011	0.0
gi 3540290 (AF057033) putative minor structural protein [Strepto...	346	2e-94
gi 2444125 (U88974) ORF46 [Streptococcus thermophilus temperate ...]	339	3e-92
gi 1934762 emb CAB07982 (Z93946) hypothetical protein [bacterio...	300	2e-80
gi 4530155 gb AAD21895.1 (AF085222) unknown [Streptococcus ther...	276	4e-73
gi 2935677 (AF032121) unknown [Streptococcus thermophilus bacter...	250	3e-65
gi 2935692 (AF032122) unknown [Streptococcus thermophilus bacter...	250	3e-65
gi 1136289 (U42597) histidine kinase A [Dictyostelium discoideum]	50	7e-05

Query= sid|114827|lan|dp1ORF006 Phage dpl ORF|45296-46987|2
(563 letters)

gi 4377165 gb AAD18987 (AE001666) SWI/SNF family helicase_2 (Ch...	171	1e-41
gi 1769947 emb CAA67095 (X98455) SNF [Bacillus cereus]	160	3e-38
gi 3329163 (AE001341) SNF/SNF family helicase [Chlamydia trachom...	159	6e-38
gi 4377149 gb AAD18973 (AE001664) SWI/SNF family helicase_1 (Ch...	157	2e-37
gi 3328995 (AE001326) SWI/SNF family helicase [Chlamydia trachom...	153	2e-36
gi 2493354 sp P75093 Y018_MYCPN HYPOTHETICAL HELICASE MG018/MG01...	146	4e-34
gi 1653748 dbj BAA18659 (D90916) helicase of the snf2/rad54 fam...	143	3e-33
gi 1763712 emb CAB05939 (Z83337) member of the SNF2 helicase fa...	143	4e-33
gi 2636153 emb CAB15645.1 (Z99122) similar to SNF2 helicase (Ba...	143	4e-33
gi 2909552 emb CAA17284 (AL021924) halZ [Mycobacterium tubercul...	140	2e-32
gi 3844627 (U39681) ATP-dependent RNA helicase, putative [Mycopla...	136	3e-31
gi 1351463 sp P47264 Y018_MYCGE HYPOTHETICAL HELICASE MG018	136	4e-31
gi 2660669 (AC002342) human Mi-2 autoantigen-like protein [Arabi...	131	2e-29
gi 1361537 pir I64201 helicase (mot1) homolog - Mycoplasma geni...	129	4e-29
gi 3482977 emb CAA20533.1 (AL031369) putative protein [Arabidop...	128	9e-29
gi 3298562 (U91543) zinc-finger helicase [Homo sapiens]	120	2e-26
gi 3875971 emb CAB02491 (Z80344) similar to helicase; cDNA EST ...	120	2e-26
gi 4557451 ref NP_001263.1 PCHD3 chromodomain helicase DNA bind...	120	2e-26
gi 2645435 (AF007780) CHD3 [Drosophila melanogaster]	118	1e-25
gi 3875165 emb CAA91798 (Z67881) Similarity to Mouse Chromodoma...	118	1e-25

Query= sid|114828|lan|dp1ORF007 Phage dpl ORF|22230-23621|3
(463 letters)

gi 2444105 (U88974) ORF26 [Streptococcus thermophilus temperate ...]	89	7e-17
gi 3318666 (U19754) BBA31 homolog [Borrelia burgdorferi]	59	7e-08
gi 2690260 (AE000790) conserved hypothetical protein [Borrelia b...	56	5e-07

Query= sid|114829|lan|dp1ORF008 Phage dpl ORF|49624-50961|1
(445 letters)

gi 4406210 gb AAD19901 (AF100420) DnaB replication fork helicase...	68	2e-10
gi 3121983 sp O25916 DNAB_HELPY REPLICATIVE DNA HELICASE >gi 231...	67	2e-10
gi 4416322 gb AAD20314 (AF106032) replicative helicase; DnaB (B...	65	9e-10
gi 4155895 (AE001551) REPLICATIVE DNA HELICASE [Helicobacter pyl...	60	4e-08
gi 3322317 (AE001191) replicative DNA helicase (dnaB) [Treponema...	58	1e-07
gi 138031 sp P04530 VG41_BPT4 PRIMASE-HELICASE (PROTEIN GP41) >g...	53	3e-06
gi 2983861 (AE000742) replicative DNA helicase [Aquifex aeolicus]	51	1e-05

Query= sid|114831|lan|dp1ORF010 Phage dpl ORF|8699-9859|2
(386 letters)

gi 2760912 (AF037258) RecA protein [Chlorobium tepidum]	133	2e-30
gi 3219851 sp P94666 RECA_CLOPE RECA PROTEIN >gi 1698591 (U61497...	129	3e-29
gi 1350566 sp P48295 RECA_STRVL RECA PROTEIN >gi 508860 (U04837)...	128	7e-29
gi 744163 prf 2014250A recA-like protein [Streptomyces violaceus]	126	3e-28
gi 730487 sp P41054 RECA_STRAM RECA PROTEIN >gi 511133 emb CAA82...	125	4e-28
gi 2687334 emb CAA15875 (AL020958) RecA protein [Streptomyces c...	125	6e-28
gi 1350565 sp P48294 RECA_STRLI RECA PROTEIN >gi 481482 pir S38...	125	6e-28

412

gi|464599|sp|P33542|RECA_AQUY RECA PROTEIN >gi|1086167|pir||A55... 123 2e-27
 gi|417636|sp|P32725|RECA_RHOSH RECA PROTEIN >gi|541307|pir||S419... 123 2e-27
 gi|2984348 (AE000775) recombination protein RecA [Aquifex aeolicus] 123 2e-27
 gi|3219854|sp|P95846|RECA_STRRM RECA PROTEIN >gi|1729800|emb|CAA... 122 4e-27
 gi|2500086|sp|Q59560|RECA_MYCSM RECA PROTEIN >gi|1430892|emb|CAA... 122 4e-27
 gi|1350567|sp|P48296|RECA_THSAQ RECA PROTEIN >gi|1072963|pir||A5... 122 6e-27
 gi|625663|pir||JX0292|recA protein - Thermus aquaticus (strain HB8) 121 1e-26
 gi|1172880|sp|P42440|RECA_CAMJE RECA PROTEIN >gi|2119991|pir||I4... 120 2e-26
 gi|4154654 (AE001453) RECA PROTEIN. [Helicobacter pylori J99] 120 2e-26
 gi|1072968|pir||C55020|recA protein - Thermus sp >gi|458472|dbj|... 120 2e-26
 gi|3219852|sp|P95469|RECA_PARDE RECA PROTEIN >gi|1825468 (U59631... 119 3e-26
 gi|2507284|sp|P42445|RECA_HELPY RECA PROTEIN >gi|2313235|gb|AAD0... 119 4e-26
 gi|1172890|sp|Q02350|RECA_STAAU RECA PROTEIN >gi|463285 (L25893)... 118 5e-26
 gi|4416209|gb|AAD20261| (AF094756) RecA protein [Bifidobacterium... 118 5e-26
 gi|2500084|sp|Q59180|RECA_BORBU RECA PROTEIN >gi|1276443 (U23457... 118 5e-26

Query= sid|114832|lan|dp1ORF011 Phage dp1 ORF|28017-29096|3
 (359 letters)

gi|2444110 (U88974) ORF31 [Streptococcus thermophilus temperate ... 187 1e-46
 gi|3320438 (AF057033) gp348 [Streptococcus thermophilus bacterio... 179 2e-44
 gi|479514|pir||S34244|hypothetical protein p38 - actinophage VWB... 62 8e-09

Query= sid|114834|lan|dp1ORF013 Phage dp1 ORF|10215-11240|3
 (341 letters)

gi|580855|emb|CAA29958| (X06803) dnaZX-like ORF put. DNA polymer... 182 2e-45
 gi|118807|sp|P09122|DP3X_BACSU DNA POLYMERASE III SUBUNITS GAMMA... 182 2e-45
 gi|98292|pir||S13786|DNA-directed DNA polymerase (EC 2.7.7.7) II... 182 2e-45
 gi|1527142 (U66040) DNA polymerase III gamma subunit [Salmonella... 172 4e-42
 gi|2494197|sp|P74876|DP3X_SALTY DNA POLYMERASE III SUBUNITS GAMM... 172 4e-42
 gi|118808|sp|P06710|DP3X_ECOLI DNA POLYMERASE III SUBUNITS GAMMA... 170 1e-41
 gi|4155207 (AE001497) DNA POLYMERASE III SUBUNITS GAMMA AND TAU ... 169 2e-41
 gi|2313841|gb|AAD07767.1| (AE000584) DNA polymerase III gamma an... 168 4e-41
 gi|2583049 (AF025391) DNA polymerase III holoenzyme tau subunit ... 166 3e-40
 gi|2984127 (AE000759) DNA polymerase III gamma subunit [Aquifex ... 166 3e-40
 gi|3861390|emb|CAA15289| (AJ235273) DNA POLYMERASE III SUBUNITS ... 165 5e-40
 gi|1169397|sp|P43746|DP3X_HABIN DNA POLYMERASE III SUBUNITS GAMM... 156 2e-37
 gi|1293572 (U49738) DNA polymerase III tau homolog DnaX [Cauloba... 151 8e-36
 gi|3328753 (AE001306) DNA Pol III Gamma and Tau [Chlamydia trach... 148 4e-35
 gi|4376294|gb|AAD18193| (AE001589) DNA Polymerase III Gamma and ... 148 5e-35
 gi|581255|emb|CAA28175| (X04487) alternate dnaZX protein (AA 1-6... 146 3e-34
 gi|2688379 (AE001151) DNA polymerase III, subunits gamma and tau... 140 2e-32
 gi|3323329 (AE001268) DNA polymerase III, subunits gamma and tau... 137 1e-31

Query= sid|114835|lan|dp1ORF014 Phage dp1 ORF|50961-51974|3
 (337 letters)

gi|1346796|sp|P47492|PRIM_MYCGE DNA PRIMASE >gi|1361496|pir||F64... 57 2e-07
 gi|740008|prf||2004290A primase [Haemophilus influenzae] 51 1e-05
 gi|1172619|sp|Q08346|PRIM_HABIN DNA PRIMASE >gi|1074033|pir||A64... 51 1e-05
 gi|1709769|sp|Q04505|PRIM_LACLA DNA PRIMASE >gi|1075726|pir||JC2... 51 1e-05
 gi|639846|dbj|BAA03516| (D14690) DNA primase [Lactococcus lactis] 51 1e-05

Query= sid|114837|lan|dp1ORF016 Phage dp1 ORF|43413-44303|3
 (296 letters)

gi|1934766|emb|CAB07986| (Z93946) N-acetylmuramoyl-L-alanine ami... 661 0.0
 gi|113676|sp|P06653|ALYS_STRPN AUTOLYSIN (N-ACETYLMURAMOYL-L-ALA... 221 4e-57
 gi|282326|pir||A42935|N-acetylmuramoyl-L-alanine amidase (EC 3.5... 219 3e-56
 gi|416618|sp|P32762|ALYS_BPHB3 LYTIC AMIDASE (N-ACETYLMURAMOYL-L... 212 2e-54
 gi|285273|pir||A42936|N-acetylmuramoyl-L-alanine amidase (EC 3.5... 212 2e-54
 gi|127787|sp|P15057|LYCA_BPCP1 LYSOZYME (ENDOLYSIN) (MURAMIDASE)... 162 4e-39
 gi|67761|pir||MUBPCP N-acetylmuramoyl-L-alanine amidase (EC 3.5... 162 4e-39
 gi|127789|sp|P19386|LYCA_BPCP9 LYSOZYME (ENDOLYSIN) (MURAMIDASE)... 160 1e-38
 gi|928832 (L44593) ORF259; putative [Lactococcus lactis phage BK... 119 2e-26
 gi|2511705|emb|CAA71783| (Y10818) sigA binding protein [Streptoc... 111 9e-24
 gi|4097980 (U72655) surface protein C [Streptococcus pneumoniae] 107 1e-22
 gi|2351768 (U89711) PspA [Streptococcus pneumoniae] 105 4e-22
 gi|2425109 (AF019904) choline binding protein A [Streptococcus p... 104 6e-22
 gi|282335|pir||A41971|surface protein pspA precursor - Streptoco... 104 1e-21
 gi|2576331|emb|CAA05158| (AJ002054) SpA protein [Streptococcus ... 103 2e-21
 gi|2127295|pir||S57962|cspC protein - Clostridium acetobutylicum... 85 6e-16
 gi|2576333|emb|CAA05159| (AJ002055) SpA protein [Streptococcus ... 84 1e-15
 gi|4106522|gb|AAD02874.1| (AF097909) excreted protein FibB [Pept... 83 3e-15
 gi|1361406|pir||S57714|cspB protein - Clostridium acetobutylicum... 82 4e-15
 gi|1914872|emb|CAB04758| (Z82001) PCPA [Streptococcus pneumoniae] 81 9e-15

413

gi 3168594 dbj BAA28613	(AB012763) SpaA (Erysipelothrix rhusiopathiae)	81	1e-14
gi 2292750 emb CAA64942	(X95646) homology to orf259 of lactococcus	80	3e-14
gi 2935696 AF032122	putative lysin (Streptococcus thermophilus)	80	3e-14
gi 4586910 dbj BAA76540.1	(AB017447) protective antigen SpaA.1	80	3e-14
gi 3540294 AF057033	lysine (Streptococcus thermophilus bacteriophage)	79	5e-14

Query= sid|114841|lan|dp1ORF020 Phage dpl ORF|1864-2658|1
(264 letters)

gi 2633745 emb CAB13247	(Z99111) similar to coenzyme PQQ synthetase	217	5e-56
gi 2808502 emb CAA12532	(AJ225561) ExsD protein (Sinorhizobium meliloti)	163	1e-39
gi 3861151 emb CAA15051	(AJ235272) unknown (Rickettsia prowazekii)	82	6e-15
gi 1652793 dbj BAA17712	(D90908) hypothetical protein (Synechococcus sp.)	76	3e-13
gi 1723815 sp P55139	YGCFF_ECOLI HYPOTHETICAL 25.0 KD PROTEIN IN	70	2e-11
gi 2984272 AE000769	hypothetical protein (Aquifex aeolicus)	66	4e-10
gi 4155435 AE001516	putative (Helicobacter pylori J99)	57	1e-07
gi 2127833 pir C64505	coenzyme PQQ synthesis protein III homolog	55	5e-07
gi 2622338 AE000890	coenzyme PQQ synthesis protein III (Methanobacterium thermoautotrophicum)	54	9e-07
gi 3257042 dbj BAA29725	(AP000003) 254aa long hypothetical protein	53	2e-06
gi 2314068 gb AAD07976.1	(AE000602) conserved hypothetical protein	52	6e-06
gi 1723816 sp P45097	YGCFF_HAEIN HYPOTHETICAL PROTEIN H11189	50	2e-05

Query= sid|114842|lan|dp1ORF021 Phage dpl ORF|2504-3295|2
(263 letters)

gi 127481 sp P19465	GCH1_BACSU GTP CYCLOHYDROLASE I (GTP-CH-I)	208	4e-53
gi 3242315 emb CAA04237	(AJ000685) GTP cyclohydrolase (Streptococcus pyogenes)	191	4e-48
gi 2494695 sp Q54769	GCH1_SYNP7 GTP CYCLOHYDROLASE I (GTP-CH-I)	189	2e-47
gi 255061 bbs 112832	(S44049) GTP cyclohydrolase I (clone hGCH-1)	187	7e-47
gi 4503949 ref NP_000152.1	PGCH1 GTP cyclohydrolase 1 (dopa-resistance)	187	7e-47
gi 2113967 emb CAB08935	(Z95557) folE (Mycobacterium tuberculosis)	187	7e-47
gi 1730240 sp P50141	GCH1_CHICK GTP CYCLOHYDROLASE I (GTP-CH-I)	185	3e-46
gi 2494696 sp Q55759	GCH1_SYNP3 GTP CYCLOHYDROLASE I (GTP-CH-I)	184	5e-46
gi 121061 sp P22288	GCH1_RAT GTP CYCLOHYDROLASE I PRECURSOR (GTP-CH-I)	184	6e-46
gi 3183014 sp Q13774	GCH1_SCHPO GTP CYCLOHYDROLASE I (GTP-CH-I)	184	6e-46
gi 3097224 emb CAA18795	(AL023093) GTP cyclohydrolase I (Mycobacterium tuberculosis)	182	2e-45
gi 2494697 sp Q19980	GCH1_CAEEL PROBABLE GTP CYCLOHYDROLASE I (GTP-CH-I)	182	2e-45
gi 462167 sp Q05915	GCH1_MOUSE GTP CYCLOHYDROLASE I PRECURSOR (GTP-CH-I)	180	7e-45
gi 1669664 emb CAA89808	(Z49706) GTP cyclohydrolase I (Dictyostelium discoideum)	180	1e-44
gi 2981082 AF052048	GTP-cyclohydrolase (Ostertagia ostertagi)	178	3e-44
gi 31954 emb CAA78908	(Z16418) GTP cyclohydrolase I (Homo sapiens)	177	8e-44
gi 551344 bbs 150280	(S71373) GTP cyclohydrolase I (mice, Peptidococcus)	174	5e-43
gi 1730247 sp P51601	GCH1_YEAST GTP CYCLOHYDROLASE I (GTP-CH-I)	174	7e-43
gi 1246912 emb CAA87397	(Z47201) GTP cyclohydrolase 1 (Saccharomyces cerevisiae)	172	2e-42
gi 1730246 sp P51595	GCH1_STRPN GTP CYCLOHYDROLASE I (GTP-CH-I)	168	3e-41
gi 2982951 AE000680	GTP cyclohydrolase I (Aquifex aeolicus)	164	6e-40

Query= sid|114843|lan|dp1ORF022 Phage dpl ORF|30896-31675|2
(259 letters)

gi 2347102 U77367	internalin (Listeria monocytogenes)	55	5e-07
gi 3123226 sp P25146	INLA_LISMO INTERNALIN A PRECURSOR	52	4e-06
gi 149674 M67471	internalin (Listeria monocytogenes)	52	4e-06

Query= sid|114850|lan|dp1ORF029 Phage dpl ORF|662-1348|2
(228 letters)

gi 2650185 AE001074	succinoglycan biosynthesis regulator (exsB)	119	2e-26
gi 3861231 emb CAA15131	(AJ235272) unknown (Rickettsia prowazekii)	117	8e-26
gi 2622210 AE000881	conserved protein (Methanobacterium thermoautotrophicum)	108	4e-23
gi 2983380 AE000709	trans-regulatory protein ExsB (Aquifex aeolicus)	88	6e-17
gi 1001327 dbj BAA10814	(D64006) ExsB (Synechocystis sp.)	88	6e-17
gi 2128055 pir B64468	hypothetical protein homolog MJ1347 - Met. (Mycobacterium tuberculosis)	83	1e-15
gi 4155143 AE001491	putative (Helicobacter pylori J99)	82	4e-15
gi 2313760 gb AAD07701.1	(AE000578) conserved hypothetical protein	80	2e-14
gi 2120814 pir S60183	protein ExsB - Rhizobium meliloti	76	3e-13
gi 2633743 emb CAB13245	(Z99111) similar to hypothetical protein	75	5e-13
gi 1175543 sp P44124	YBAX_HAEIN HYPOTHETICAL PROTEIN H11191	74	1e-12
gi 2495537 sp P77756	YBAX_ECOLI HYPOTHETICAL 25.5 KD PROTEIN IN	71	5e-12
gi 3256471 dbj BAA29154.1	(AP000001) 269aa long hypothetical protein	67	1e-10
gi 2921156 AF022216	aluminum resistance protein (Arthrobacter)	54	1e-06

Query= sid|114855|lan|dp1ORF034 Phage dpl ORF|131-652|2
(173 letters)

gi 2633746 emb CAB13248	(Z99111) similar to hypothetical protein	220	4e-57
-------------------------	--	-----	-------

414

gi|4155926 (AE001554) putative (Helicobacter pylori J99) 162 1e-39
gi|2314588|gb|AAD08456.1| (AE000642) conserved hypothetical prot... 161 3e-39
gi|2983458 (AE000714) hypothetical protein (Aquifex aeolicus) 103 9e-22
gi|1006604|dbj|BAA10757| (D64005) hypothetical protein (Synechoc... 87 6e-17
gi|2967529 (U11045) unknown (Buchnera aphidicola) 79 2e-14
gi|2495654|sp|Q46920|YQCD_ECOLI HYPOTHETICAL 32.6 KD PROTEIN IN ... 69 2e-11
gi|1175604|sp|P44153|YQCD_HAEIN HYPOTHETICAL PROTEIN HI1291 >gi|... 63 1e-09
gi|3860642|emb|CAA14543| (AJ235270) unknown (Rickettsia prowazekii) 56 1e-07

Query= sid|114857|lan|dp1ORF036 Phage dp1 ORF|48808-49362|1
(184 letters)

gi|1353529 (U38906) ORF12 (Bacteriophage rlt) 53 1e-06

Query= sid|114859|lan|dp1ORF038 Phage dp1 ORF|1350-1871|3
(173 letters)

gi|1175542|sp|P44123|YB90_HAEIN HYPOTHETICAL PROTEIN HI1190 >gi|... 100 6e-21
gi|2982977 (AE000681) hypothetical protein (Aquifex aeolicus) 67 7e-11
gi|3860744|emb|CAA14645| (AJ235270) unknown (Rickettsia prowazekii) 65 3e-10
gi|2650193 (AE001074) conserved hypothetical protein (Archaeoglo... 58 4e-08
gi|3258383|dbj|BAA31066.1| (AP000007) 157aa long hypothetical pr... 55 2e-07
gi|1001713|dbj|BAA10550| (D64004) hypothetical protein (Synechoc... 50 8e-06
gi|4155434 (AE001516) putative (Helicobacter pylori J99) 50 1e-05

Query= sid|114860|lan|dp1ORF039 Phage dp1 ORF|3306-3803|3
(165 letters)

gi|1922884|emb|CAA68244| (X99978) ORF7; hydrophobic protein (Lact... 64 5e-10

Query= sid|114862|lan|dp1ORF041 Phage dp1 ORF|8208-8699|3
(163 letters)

gi|2522313 (AF012906) dUTPase homolog (Bacillus subtilis) >gi|26... 108 2e-23
gi|2634150|emb|CAB13650| (Z99113) similar to deoxyuridine 5'-tri... 108 3e-23
gi|3913546|sp|O54134|DUT_STRCO DEOXYURIDINE 5'-TRIPHOSPHATE NUCL... 56 2e-07
gi|3913542|sp|O48500|DUT_BPTS DEOXYURIDINE 5'-TRIPHOSPHATE NUCL... 52 3e-06
gi|3913548|sp|O68992|DUT_CHLTE DEOXYURIDINE 5'-TRIPHOSPHATE NUCL... 50 1e-05

Query= sid|114867|lan|dp1ORF046 Phage dp1 ORF|42774-43202|3
(142 letters)

gi|1934764|emb|CAB07984| (Z93946) hypothetical protein (bacterio... 287 2e-77

Query= sid|114901|lan|dp1ORF080 Phage dp1 ORF|42490-42759|1
(89 letters)

gi|1934763|emb|CAB07983| (Z93946) hypothetical protein (bacterio... 147 1e-35

Query= sid|114912|lan|dp1ORF091 Phage dp1 ORF|43189-43413|1
(74 letters)

gi|1934765|emb|CAB07985| (Z93946) holin (bacteriophage Dp-1) 63 2e-10

Table 32

Sequence of Dp1 published by Sheehan and al., 4731 nucleotides.

1	ttttaaatttt	ttgacaaagt	taattcaaat	tgtaccgctg	aagcaatttt	ccatgtatcc	actcaaagtt
71	gttcagtggtg	gctcaatcat	attaaaaatc	aacttggtaa	tatctctact	cccttttagtg	aagcagagga
141	agaccttaaa	tatcgaattg	actcaaaagc	cgatcaaaag	ctaactaacc	aacagttgac	ggcactcacg
211	gaaaaggctc	aactacatga	cgcagaactg	aaagctaagg	ctacaatgga	gcagtttaagt	aacttagaaa
281	aggcttatga	aggtagaatg	aaagctaagt	aaagagctat	caacaaatcg	gaaccgcacc	taactcttagc
351	ggcaagtctga	attgaagcta	ctatccaaga	acttggcggg	ctacggggaac	tgaagaagtt	cgtcgacagt
421	tgcattgagct	cttctaatac	aggcttaatt	atcggttaaga	acgacggtag	ctctaccatt	aaggratcaa
491	gtgaccgaat	ttctatgttc	tccgcaggga	atgaagttat	gtaccttacg	caagggttca	ttcacatcga
561	taacggggtc	tttaccat	ccattccaag	cggccgattt	agaacgggaac	aatactcgtt	taatccagac
631	atgaacgtga	ttcgggtatg	aggataagga	gaataaacatg	acaaaattta	tcaactcata	cggccctctt
701	cacttgaaac	tttactgcga	acaagtttag	caggacgtaa	cgaacaactc	ctccgcgagt	agttggcgag
771	ctactgtcga	ccgcgatgga	gcttatcgaa	cgttgactta	tggaaatatt	agtaaccttt	ccgtatgggt
841	aaatgggtca	agtggttcata	gcagtcaccc	agactacgac	acgtccggcg	aagaggtaac	gctcgaaggt
911	ggagaagtga	ctgttctctca	caatagtgcac	gggacaaaga	caatgtccgt	ttgggcttcg	tttgacctca
981	acaacggcgt	tcacggaaat	atcactatct	ctactaatta	cacttttagac	agttatccaa	ggctcacaca
1051	gatttctagt	tttgagggaa	atcgaaatct	aggatcttta	catacgggta	tcttcaaccg	aaaagtgaac
1121	tctttttacgc	atcaagtttg	gtaccgagtt	ttcggtagcg	actggataga	tttaggtaag	aaccatacta
1191	ctagcgatc	ctttacgcgc	tcactggact	tagcaaggtta	cttacctaaa	tcaagttccg	gaacaatgga
1261	catctgtatt	cgaacctata	acggaaactac	gcaaatgggt	agtgacgtct	attcaaacgg	atggagggtc
1331	aacatccccg	attcagtagc	ttctactttt	tcgggcattt	ctttagtaga	cacgacttca	gcggttcgac
1401	agatttttaac	agggaaacac	ttcttccaaa	tcattgtcgaa	cattcaagtc	aacttcaaca	atgcttccgg
1471	cgcttacgga	tcactatctc	aagcatttca	cgctgagctc	gtaggtaaaa	accaaggtat	caacgaaaac
1541	ggcgggcaat	tgggtatgat	gaactttaat	ggctccgcta	ccgttaagagc	atgggttaca	gacacgcgag
1611	gaaaaacaatc	gaaacgtccaa	gacgtatcta	tcaatgttat	agaatactat	ggaccgtcta	tcaatttctc
1681	cggtcaacgt	actcgtcaaa	atcctgcaat	tatccaagct	cttcgaaatg	ctaagggtcg	acctataacg
1751	gtaggagggt	aacagaaaaa	catcatgcaa	attaccctct	ccgtggcgcc	gttgaaactc	actaatttca
1821	cagaagatag	agggtcggcg	tcagggacgt	tcactactat	ttccctactg	actaactcgt	ccgcgaactt
1891	agctggtaac	tacgggcggg	acaagttcta	catagtttaag	gctaaaatcc	aagacaggtt	cacttcgact
1961	gaatttttagt	ctacgggtacc	taccgaatca	gtagtctcta	actatgacaa	ggacgggtcg	cttggagtgt
2031	gtaaggttgt	agaacaaggg	aaggcagggg	caattgatgc	agcaggtgat	atatacgtcg	gaggtcgaca
2101	agttcaacag	tttcagctca	ctgataataa	tggagcattg	aacagggggtc	aatataacga	tgttggaaata
2171	agcgtgaaac	agagtttaca	tggcgaaagta	ggcaaacctc	ggcaaacctc	acgggaactc	gaggtgaaatg
2241	gggactatct	caaaatttct	gggttagatg	ctggaaaaatg	gttcaatcct	tcattacaat	gtcaggaaaga
2311	atgttcatca	ggacagcgaa	cgatggaaac	agctggagac	ctaacaagtg	gaaagaggtt	ctattttaagc
2381	aagacttcga	acagaataat	tggcgaaaac	ttgttcttca	aagtggttg	aaccatcact	caacctatgg
2451	cgaacgattc	tattcgaaaa	ctcttgacgg	catagtatat	ttgagaggaa	atgtgcataa	aggacttatc
2521	gacaaagagg	ctactattgc	agtacttctt	gaaggattta	gaacgaaagt	ttcaatgtat	cttcaggctc
2591	tcaataaactc	atatggaaat	gccatttctat	gtatatcac	tgacggaaaga	cttgtggtga	aatcgaatgt
2661	agataaattct	tgggttaaat	tagacaatgt	ctcatttctg	attttaatttg	agctgaaatc	atgttatcaat
2731	atttttttaga	aaggagggtga	gaactatggt	gaaccttaca	aaatcgcgcc	aaattgtggc	agagttcact
2801	attggacaag	gagctgaaaa	gaaacttctc	aaaacaacga	ttgtgaacat	tgtatgcaac	gcagtatcaa
2871	ccgtctctga	aactcttcat	gacccagact	tgtatgtctg	gaaccgtcga	gaacttcgag	ctgacgagca
2941	aaaactctcg	gaaactctgt	acgcaatcga	agatgaaatt	aatagctgga	gcgggggaaa	aaagggggag
3011	cccggtctca	acaggctgaa	taaggaggcg	tcaatctatg	ccaatgtggc	taaacgacac	cgcagctctg
3081	acgacgatta	ttacagcggt	cagcgagggt	cttactgtcc	tactaaataa	gttattcgaa	tggaaatcga
3151	ataaagccaa	gagcgtttta	gaggatatct	ctacaactct	tagcactctt	aaacagcagg	tcgacgggat
3221	tgaccacaa	acagtagcaa	tcaatcacca	aatgacgtc	attcaagacg	gaactagaaa	aatccaactg
3291	taccgtcttt	atcacgactt	aaaaaggga	gtgataaacg	gctatacaac	tctcgaccat	tttagagagc
3361	tctctatttt	attcgaaagt	tataagaacc	ttggcggaaa	tgggtgaagt	gaagccttct	atgaaaaata
3431	caagaaatta	ccaattaggg	aggaagattt	agatgaaact	atctaacgaa	caatatgacg	tagcaaaaga
3501	cgtggtaaac	gtagtctgtc	cagcagcgat	tgacttaatt	acaggtcttg	gagcgttgta	tcaatttgac
3571	actactgcta	tcacaggaac	cattgcactt	cttgcaactt	ttgcaggtac	tgttctagga	gtttctagcc
3641	gaactacca	aaaggaaaca	gaagctcaaa	acaatgaggt	ggaataatgg	gagtcgatat	tgaaaaaggc
3711	gttgcgtgga	tgacggcccg	aaagggtcga	gtactttata	gcactggactt	tcgagacggg	cctgatagct
3781	atgactgctc	aagttctatg	tactatgctc	tccgctcagc	cggagcttca	agtgtctggt	gggcagctcaa
3851	tactgagtag	atgcacgcac	ggcttatatg	aaacgggtat	gaactaatta	gtgaaaatgc	tcgctgggat
3921	gctaaacgag	gcgacatctt	catctgggga	cgcaaaagtg	ctagcgcagg	cgtctggagg	catcacggga
3991	tgttcattga	cagtgataac	atcattcact	gcaactacgc	ctacgacgga	atttccgcta	acgaccacga
4061	tgagcgttgg	tactatgcag	gtcaacctta	ctactacgtc	tatcgcttga	ctaacgcaa	tgctcaaccg
4131	gctgagaga	aaacttggctg	gcagaaagat	gctactgggt	tctggttacg	tcgagcaaac	ggaaacttatc
4201	caaaagatga	gttcgagtat	atcgaaagaa	acaagctctg	gttctacttt	gacgaccaag	gctacatgct
4271	cgctgagaaa	tgggtgaaac	atactgatgg	aaattgggtat	tggttcgacc	gtgacggata	catggctacg
4341	tcattggaaac	ggattggcga	gtcatggtag	tacttcaatc	gcgattggtc	aatggtaacc	gggttgattt
4411	agatttaaga	taattgggtat	tattgtgatg	ctaccaacgg	cgacatgaaa	tcgaatcgct	ttatccgtta
4481	taacgacggc	tgggtatctac	tattacggga	cggacgtctg	gcagataaac	ctcaattcac	cgtagagccg
4551	gacgggctca	ttactgctaa	agcttaaaat	atagagagga	ggaagctctt	ttcttaatat	tgtttctctt
4621	aatcccgcaa	gggttcgacc	ctgcgggggt	tatgtgtcgt	gaattactct	atttacttat	tcgaagattt
4691	caattataat	taaatatca	acgagattca	taattggagg	aatg		

Table 33

Streptococcus accession numbers

gi 5776553 gb AF026471.2 AF026471 [5776553]	gi 5231200 gb AF157824.1 AF157824 [5231200]
gi 5410470 gb AF139890.1 AF139890 [5410470]	gi 5231197 gb AF157823.1 AF157823 [5231197]
gi 5410468 gb AF139889.1 AF139889 [5410468]	gi 5231194 gb AF157822.1 AF157822 [5231194]
gi 5410466 gb AF139888.1 AF139888 [5410466]	gi 5231191 gb AF157821.1 AF157821 [5231191]
gi 5410464 gb AF139887.1 AF139887 [5410464]	gi 5231188 gb AF157820.1 AF157820 [5231188]
gi 5410462 gb AF139886.1 AF139886 [5410462]	gi 5231185 gb AF157819.1 AF157819 [5231185]
gi 5410460 gb AF139885.1 AF139885 [5410460]	gi 5231182 gb AF157818.1 AF157818 [5231182]
gi 5410458 gb AF139884.1 AF139884 [5410458]	gi 5231179 gb AF157817.1 AF157817 [5231179]
gi 5410456 gb AF139883.1 AF139883 [5410456]	gi 4336851 gb AF106138.1 AF106138 [4336851]
gi 3093394 emb AJ005697.1 SPN5697 [3093394]	gi 4336848 gb AF106137.1 AF106137 [4336848]
gi 5759208 gb AF171873.1 AF171873 [5759208]	gi 4336845 gb AF106136.1 AF106136 [4336845]
gi 5758311 gb AF162664.1 AF162664 [5758311]	gi 4336842 gb AF106135.1 AF106135 [4336842]
gi 5739313 gb AF161701.1 AF161701 [5739313]	gi 4336839 gb AF106134.1 AF106134 [4336839]
gi 5739310 gb AF161700.1 AF161700 [5739310]	gi 4336836 gb AF106133.1 AF106133 [4336836]
gi 5726354 gb AF159448.1 AF159448 [5726354]	gi 4336833 gb AF106132.1 AF106132 [4336833]
gi 5726290 gb AF127143.1 AF127143 [5726290]	gi 3907597 gb AF094575.1 AF094575 [3907597]
gi 5712666 gb AF140784.1 AF140784 [5712666]	gi 5030425 gb AF061748.2 AF061748 [5030425]
gi 4218525 emb AJ009639.1 SPAJ9639 [4218525]	gi 4902881 emb AJ239004.1 SPN239004 [4902881]
gi 5616524 gb AF169483.1 AF169483 [5616524]	gi 5001710 gb AF112358.1 AF112358 [5001710]
gi 5579395 gb AF162656.1 AF162656 [5579395]	gi 5001690 gb AF106539.1 AF106539 [5001690]
gi 5579393 gb AF162655.1 AF162655 [5579393]	gi 4973271 gb AF144420.1 AF144420 [4973271]
gi 5578890 emb AJ131985.1 SPN131985 [5578890]	gi 4973269 gb AF144419.1 AF144419 [4973269]
gi 5566442 gb AF167442.1 AF167442 [5566442]	gi 4973267 gb AF144418.1 AF144418 [4973267]
gi 5459332 emb AJ243540.1 EVE243540 [5459332]	gi 4928190 gb AF129757.1 AF129757 [4928190]
gi 5305398 gb AF072811.1 AF072811 [5305398]	gi 4927743 gb AF126061.1 AF126061 [4927743]
gi 5295921 emb AJ242698.1 SPN242698 [5295921]	gi 4927742 gb AF126060.1 AF126060 [4927742]
gi 5295920 emb AJ242697.1 SPN242697 [5295920]	gi 4927741 gb AF126059.1 AF126059 [4927741]
gi 5295919 emb AJ242696.1 SPN242696 [5295919]	gi 4495247 emb AJ240675.1 SPN240675 [4495247]
gi 5295918 emb AJ242695.1 SPN242695 [5295918]	gi 4495245 emb AJ240670.1 SPN240670 [4495245]
gi 4583522 gb AF140356.1 AF140356 [4583522]	gi 4495243 emb AJ240669.1 SPN240669 [4495243]
gi 5231206 gb AF157826.1 AF157826 [5231206]	gi 4495241 emb AJ240668.1 SPN240668 [4495241]
gi 5231203 gb AF157825.1 AF157825 [5231203]	gi 4495239 emb AJ240667.1 SPN240667 [4495239]

gi|4495237|emb|AJ240666.1|SPN240666
[4495237]
gi|4495235|emb|AJ240665.1|SPN240665
[4495235]
gi|4495233|emb|AJ240664.1|SPN240664
[4495233]
gi|4495231|emb|AJ240663.1|SPN240663
[4495231]
gi|4495229|emb|AJ240662.1|SPN240662
[4495229]
gi|4495227|emb|AJ240661.1|SPN240661
[4495227]
gi|4495225|emb|AJ240660.1|SPN240660
[4495225]
gi|4495223|emb|AJ240659.1|SPN240659
[4495223]
gi|4495221|emb|AJ240658.1|SPN240658
[4495221]
gi|4495219|emb|AJ240657.1|SPN240657
[4495219]
gi|4495217|emb|AJ240656.1|SPN240656
[4495217]
gi|4495215|emb|AJ240655.1|SPN240655
[4495215]
gi|4495213|emb|AJ240654.1|SPN240654
[4495213]
gi|4495211|emb|AJ240653.1|SPN240653
[4495211]
gi|4495209|emb|AJ240652.1|SPN240652
[4495209]
gi|4495207|emb|AJ240651.1|SPN240651
[4495207]
gi|4495205|emb|AJ240650.1|SPN240650
[4495205]
gi|4495203|emb|AJ240649.1|SPN240649
[4495203]
gi|4495201|emb|AJ240648.1|SPN240648
[4495201]
gi|4495199|emb|AJ240647.1|SPN240647
[4495199]
gi|4495197|emb|AJ240644.1|SPN240644
[4495197]
gi|4495195|emb|AJ240643.1|SPN240643
[4495195]
gi|4495193|emb|AJ240642.1|SPN240642
[4495193]
gi|4495191|emb|AJ240641.1|SPN240641
[4495191]

gi|4495189|emb|AJ240640.1|SPN240640
[4495189]
gi|4495187|emb|AJ240639.1|SPN240639
[4495187]
gi|4495185|emb|AJ240638.1|SPN240638
[4495185]
gi|4495183|emb|AJ240637.1|SPN240637
[4495183]
gi|4495181|emb|AJ240636.1|SPN240636
[4495181]
gi|4495179|emb|AJ240635.1|SPN240635
[4495179]
gi|4495177|emb|AJ240634.1|SPN240634
[4495177]
gi|4495175|emb|AJ240633.1|SPN240633
[4495175]
gi|4495173|emb|AJ240630.1|SPN240630
[4495173]
gi|4495171|emb|AJ240629.1|SPN240629
[4495171]
gi|4495169|emb|AJ240628.1|SPN240628
[4495169]
gi|4495167|emb|AJ240627.1|SPN240627
[4495167]
gi|4495165|emb|AJ240626.1|SPN240626
[4495165]
gi|4495163|emb|AJ240625.1|SPN240625
[4495163]
gi|4495161|emb|AJ240624.1|SPN240624
[4495161]
gi|4495159|emb|AJ240623.1|SPN240623
[4495159]
gi|4495157|emb|AJ240622.1|SPN240622
[4495157]
gi|4495155|emb|AJ240621.1|SPN240621
[4495155]
gi|4495153|emb|AJ240620.1|SPN240620
[4495153]
gi|4495151|emb|AJ240619.1|SPN240619
[4495151]
gi|4495149|emb|AJ240616.1|SPN240616
[4495149]
gi|4495147|emb|AJ240615.1|SPN240615
[4495147]
gi|4495145|emb|AJ240614.1|SPN240614
[4495145]
gi|4495143|emb|AJ240613.1|SPN240613
[4495143]

gi|4495141|emb|AJ240612.1|SPN240612
[4495141]
gi|4495139|emb|AJ240611.1|SPN240611
[4495139]
gi|4495137|emb|AJ240610.1|SPN240610
[4495137]
gi|4495135|emb|AJ240609.1|SPN240609
[4495135]
gi|4495133|emb|AJ240608.1|SPN240608
[4495133]
gi|4495131|emb|AJ240607.1|SPN240607
[4495131]
gi|4495129|emb|AJ240606.1|SPN240606
[4495129]
gi|4883698|gb|AF079807.1|AF079807 [4883698]
gi|4838562|gb|AF145055.1|AF145055 [4838562]
gi|4063727|gb|L29324.1|STRINTE [4063727]
gi|3093401|emb|AJ005619.1|SPAJS619 [3093401]
gi|4103889|gb|AF029368.1|AF029368 [4103889]
gi|2897689|dbj|D63805.1|D63805 [2897689]
gi|4566771|gb|AF117741.1|AF117741 [4566771]
gi|4566768|gb|AF117740.1|AF117740 [4566768]
gi|4538836|emb|AJ240793.1|SPN240793
[4538836]
gi|4538832|emb|AJ240792.1|SPN240792
[4538832]
gi|4538828|emb|AJ240791.1|SPN240791
[4538828]
gi|4538824|emb|AJ240790.1|SPN240790
[4538824]
gi|4538821|emb|AJ240789.1|SPN240789
[4538821]
gi|4538818|emb|AJ240788.1|SPN240788
[4538818]
gi|4538815|emb|AJ240787.1|SPN240787
[4538815]
gi|4538812|emb|AJ240786.1|SPN240786
[4538812]
gi|4538809|emb|AJ240785.1|SPN240785
[4538809]
gi|4538806|emb|AJ240784.1|SPN240784
[4538806]
gi|4538803|emb|AJ240783.1|SPN240783
[4538803]
gi|4538800|emb|AJ240782.1|SPN240782
[4538800]

gi|4538797|emb|AJ240781.1|SPN240781
[4538797]
gi|4538794|emb|AJ240780.1|SPN240780
[4538794]
gi|4538791|emb|AJ240779.1|SPN240779
[4538791]
gi|4538788|emb|AJ240778.1|SPN240778
[4538788]
gi|4538785|emb|AJ240777.1|SPN240777
[4538785]
gi|4538782|emb|AJ240776.1|SPN240776
[4538782]
gi|4538779|emb|AJ240775.1|SPN240775
[4538779]
gi|4538776|emb|AJ240774.1|SPN240774
[4538776]
gi|4538773|emb|AJ240773.1|SPN240773
[4538773]
gi|4538770|emb|AJ240772.1|SPN240772
[4538770]
gi|4538767|emb|AJ240771.1|SPN240771
[4538767]
gi|4538764|emb|AJ240770.1|SPN240770
[4538764]
gi|4538761|emb|AJ240769.1|SPN240769
[4538761]
gi|4538758|emb|AJ240768.1|SPN240768
[4538758]
gi|4538755|emb|AJ240767.1|SPN240767
[4538755]
gi|4538752|emb|AJ240766.1|SPN240766
[4538752]
gi|4538749|emb|AJ240765.1|SPN240765
[4538749]
gi|4538746|emb|AJ240764.1|SPN240764
[4538746]
gi|4538743|emb|AJ240760.1|SPN240760
[4538743]
gi|4538740|emb|AJ240759.1|SPN240759
[4538740]
gi|4538737|emb|AJ240758.1|SPN240758
[4538737]
gi|4538734|emb|AJ240757.1|SPN240757
[4538734]
gi|4538731|emb|AJ240756.1|SPN240756
[4538731]
gi|4538728|emb|AJ240755.1|SPN240755
[4538728]

gi|4538725|emb|AJ240754.1|SPN240754
[4538725]
gi|4538722|emb|AJ240753.1|SPN240753
[4538722]
gi|4538719|emb|AJ240752.1|SPN240752
[4538719]
gi|4538716|emb|AJ240751.1|SPN240751
[4538716]
gi|4538713|emb|AJ240750.1|SPN240750
[4538713]
gi|4538710|emb|AJ240749.1|SPN240749
[4538710]
gi|4538707|emb|AJ240748.1|SPN240748
[4538707]
gi|4538704|emb|AJ240747.1|SPN240747
[4538704]
gi|4538701|emb|AJ240746.1|SPN240746
[4538701]
gi|4538698|emb|AJ240745.1|SPN240745
[4538698]
gi|4538695|emb|AJ240744.1|SPN240744
[4538695]
gi|4538692|emb|AJ240743.1|SPN240743
[4538692]
gi|4538689|emb|AJ240742.1|SPN240742
[4538689]
gi|4538686|emb|AJ240741.1|SPN240741
[4538686]
gi|4538683|emb|AJ240740.1|SPN240740
[4538683]
gi|4538680|emb|AJ240739.1|SPN240739
[4538680]
gi|4538677|emb|AJ240738.1|SPN240738
[4538677]
gi|4530444|gb|AF118229.1|AF118229 [4530444]
gi|4519253|dbj|AB015852.1|AB015852 [4519253]
gi|4519251|dbj|AB015851.1|AB015851 [4519251]
gi|4519249|dbj|AB015850.1|AB015850 [4519249]
gi|4519247|dbj|AB015849.1|AB015849 [4519247]
gi|4519245|dbj|AB015848.1|AB015848 [4519245]
gi|4519243|dbj|AB015847.1|AB015847 [4519243]
gi|4519241|dbj|AB015846.1|AB015846 [4519241]
gi|4519239|dbj|AB011210.1|AB011210 [4519239]
gi|4519237|dbj|AB011209.1|AB011209 [4519237]
gi|4519235|dbj|AB011208.1|AB011208 [4519235]

gi|4519233|dbj|AB011207.1|AB011207 [4519233]
gi|4519231|dbj|AB011206.1|AB011206 [4519231]
gi|4519229|dbj|AB011205.1|AB011205 [4519229]
gi|4519227|dbj|AB011204.1|AB011204 [4519227]
gi|4519225|dbj|AB011203.1|AB011203 [4519225]
gi|4519223|dbj|AB011202.1|AB011202 [4519223]
gi|4519221|dbj|AB011201.1|AB011201 [4519221]
gi|4519219|dbj|AB011200.1|AB011200 [4519219]
gi|4519217|dbj|AB011199.1|AB011199 [4519217]
gi|4519215|dbj|AB011198.1|AB011198 [4519215]
gi|4495127|emb|AJ240605.1|SPN240605
[4495127]
gi|4468031|emb|AJ132957.1|SPN132957
[4468031]
gi|4468029|emb|AJ132956.1|SPN132956
[4468029]
gi|4218532|emb|AJ010312.1|SPN010312
[4218532]
gi|4456852|emb|AJ236792.1|SPN236792
[4456852]
gi|4456850|emb|AJ236791.1|SPN236791
[4456850]
gi|4456848|emb|AJ236790.1|SPN236790
[4456848]
gi|4456846|emb|AJ236789.1|SPN236789
[4456846]
gi|3550644|emb|AJ006987.1|SPAJ6987 [3550644]
gi|3550625|emb|AJ006986.1|SPAJ6986 [3550625]
gi|4416518|gb|AF014458.2|AF014458 [4416518]
gi|4406260|gb|AF105116.1|AF105116 [4406260]
gi|4406257|gb|AF105115.1|AF105115 [4406257]
gi|4406254|gb|AF105114.1|AF105114 [4406254]
gi|4406246|gb|AF105113.1|AF105113 [4406246]
gi|4406243|gb|AF105112.1|AF105112 [4406243]
gi|4138533|emb|AJ005815.1|SPN5815 [4138533]
gi|3821726|emb|AJ232433.1|SPN232433
[3821726]
gi|3821724|emb|AJ232432.1|SPN232432
[3821724]
gi|3821722|emb|AJ232431.1|SPN232431
[3821722]
gi|3821720|emb|AJ232430.1|SPN232430
[3821720]

gi|3821718|emb|AJ232429.1|SPN232429
[3821718]
gi|3821716|emb|AJ232428.1|SPN232428
[3821716]
gi|3821714|emb|AJ232427.1|SPN232427
[3821714]
gi|3821712|emb|AJ232426.1|SPN232426
[3821712]
gi|3821710|emb|AJ232425.1|SPN232425
[3821710]
gi|3821708|emb|AJ232424.1|SPN232424
[3821708]
gi|3821706|emb|AJ232423.1|SPN232423
[3821706]
gi|3821704|emb|AJ232422.1|SPN232422
[3821704]
gi|3821702|emb|AJ232421.1|SPN232421
[3821702]
gi|3821700|emb|AJ232420.1|SPN232420
[3821700]
gi|3821698|emb|AJ232419.1|SPN232419
[3821698]
gi|3821696|emb|AJ232418.1|SPN232418
[3821696]
gi|3821694|emb|AJ232417.1|SPN232417
[3821694]
gi|3821692|emb|AJ232416.1|SPN232416
[3821692]
gi|3821690|emb|AJ232415.1|SPN232415
[3821690]
gi|3821688|emb|AJ232414.1|SPN232414
[3821688]
gi|3821686|emb|AJ232413.1|SPN232413
[3821686]
gi|3821684|emb|AJ232412.1|SPN232412
[3821684]
gi|3821682|emb|AJ232411.1|SPN232411
[3821682]
gi|3821680|emb|AJ232410.1|SPN232410
[3821680]
gi|3821678|emb|AJ232409.1|SPN232409
[3821678]
gi|3821676|emb|AJ232408.1|SPN232408
[3821676]
gi|3821674|emb|AJ232407.1|SPN232407
[3821674]
gi|3821672|emb|AJ232406.1|SPN232406
[3821672]

gi|3821670|emb|AJ232405.1|SPN232405
[3821670]
gi|3821668|emb|AJ232404.1|SPN232404
[3821668]
gi|3821666|emb|AJ232403.1|SPN232403
[3821666]
gi|3821664|emb|AJ232402.1|SPN232402
[3821664]
gi|3821662|emb|AJ232401.1|SPN232401
[3821662]
gi|3821660|emb|AJ232399.1|SPN232399
[3821660]
gi|3821658|emb|AJ232398.1|SPN232398
[3821658]
gi|3821656|emb|AJ232397.1|SPN232397
[3821656]
gi|3821654|emb|AJ232396.1|SPN232396
[3821654]
gi|3821652|emb|AJ232395.1|SPN232395
[3821652]
gi|3821650|emb|AJ232394.1|SPN232394
[3821650]
gi|3821648|emb|AJ232393.1|SPN232393
[3821648]
gi|3821646|emb|AJ232392.1|SPN232392
[3821646]
gi|3821644|emb|AJ232391.1|SPN232391
[3821644]
gi|3821642|emb|AJ232390.1|SPN232390
[3821642]
gi|3821640|emb|AJ232389.1|SPN232389
[3821640]
gi|3821638|emb|AJ232388.1|SPN232388
[3821638]
gi|3821636|emb|AJ232387.1|SPN232387
[3821636]
gi|3821634|emb|AJ232386.1|SPN232386
[3821634]
gi|3821632|emb|AJ232385.1|SPN232385
[3821632]
gi|3821630|emb|AJ232384.1|SPN232384
[3821630]
gi|3821628|emb|AJ232383.1|SPN232383
[3821628]
gi|3821626|emb|AJ232382.1|SPN232382
[3821626]
gi|3821624|emb|AJ232381.1|SPN232381
[3821624]

gi|3821622|emb|AJ232380.1|SPN232380
[3821622]
gi|3821620|emb|AJ232379.1|SPN232379
[3821620]
gi|3821618|emb|AJ232378.1|SPN232378
[3821618]
gi|3821616|emb|AJ232377.1|SPN232377
[3821616]
gi|3821614|emb|AJ232376.1|SPN232376
[3821614]
gi|3821612|emb|AJ232375.1|SPN232375
[3821612]
gi|3821610|emb|AJ232373.1|SPN232373
[3821610]
gi|3821608|emb|AJ232372.1|SPN232372
[3821608]
gi|3821606|emb|AJ232371.1|SPN232371
[3821606]
gi|3821604|emb|AJ232370.1|SPN232370
[3821604]
gi|3821602|emb|AJ232369.1|SPN232369
[3821602]
gi|3821600|emb|AJ232368.1|SPN232368
[3821600]
gi|3821598|emb|AJ232367.1|SPN232367
[3821598]
gi|3821596|emb|AJ232366.1|SPN232366
[3821596]
gi|3821594|emb|AJ232365.1|SPN232365
[3821594]
gi|3820454|emb|AJ007367.1|SPN7367 [3820454]
gi|3821592|emb|AJ232364.1|SPN232364
[3821592]
gi|3821590|emb|AJ232363.1|SPN232363
[3821590]
gi|3821588|emb|AJ232362.1|SPN232362
[3821588]
gi|3821586|emb|AJ232361.1|SPN232361
[3821586]
gi|3821584|emb|AJ232360.1|SPN232360
[3821584]
gi|3821582|emb|AJ232359.1|SPN232359
[3821582]
gi|3821580|emb|AJ232358.1|SPN232358
[3821580]
gi|3821578|emb|AJ232357.1|SPN232357
[3821578]

gi|3821576|emb|AJ232356.1|SPN232356
[3821576]
gi|3821574|emb|AJ232355.1|SPN232355
[3821574]
gi|3821572|emb|AJ232353.1|SPN232353
[3821572]
gi|3821570|emb|AJ232352.1|SPN232352
[3821570]
gi|3821568|emb|AJ232351.1|SPN232351
[3821568]
gi|3821566|emb|AJ232350.1|SPN232350
[3821566]
gi|3821564|emb|AJ232349.1|SPN232349
[3821564]
gi|3821562|emb|AJ232348.1|SPN232348
[3821562]
gi|3821560|emb|AJ232347.1|SPN232347
[3821560]
gi|3821558|emb|AJ232346.1|SPN232346
[3821558]
gi|3821556|emb|AJ232345.1|SPN232345
[3821556]
gi|3821554|emb|AJ232344.1|SPN232344
[3821554]
gi|3821552|emb|AJ232343.1|SPN232343
[3821552]
gi|3821550|emb|AJ232342.1|SPN232342
[3821550]
gi|3821548|emb|AJ232341.1|SPN232341
[3821548]
gi|3821546|emb|AJ232340.1|SPN232340
[3821546]
gi|3821544|emb|AJ232339.1|SPN232339
[3821544]
gi|3821542|emb|AJ232338.1|SPN232338
[3821542]
gi|3821540|emb|AJ232337.1|SPN232337
[3821540]
gi|3821538|emb|AJ232336.1|SPN232336
[3821538]
gi|3821536|emb|AJ232335.1|SPN232335
[3821536]
gi|3821534|emb|AJ232334.1|SPN232334
[3821534]
gi|3821532|emb|AJ232333.1|SPN232333
[3821532]
gi|3821530|emb|AJ232332.1|SPN232332
[3821530]

gi|3821528|emb|AJ232331.1|SPN232331
[3821528]
gi|3821526|emb|AJ232330.1|SPN232330
[3821526]
gi|3821524|emb|AJ232329.1|SPN232329
[3821524]
gi|3821522|emb|AJ232328.1|SPN232328
[3821522]
gi|3821520|emb|AJ232327.1|SPN232327
[3821520]
gi|3821518|emb|AJ232326.1|SPN232326
[3821518]
gi|3821516|emb|AJ232325.1|SPN232325
[3821516]
gi|3821514|emb|AJ232324.1|SPN232324
[3821514]
gi|3821512|emb|AJ232322.1|SPN232322
[3821512]
gi|3821510|emb|AJ232321.1|SPN232321
[3821510]
gi|3821508|emb|AJ232320.1|SPN232320
[3821508]
gi|3821506|emb|AJ232319.1|SPN232319
[3821506]
gi|3821504|emb|AJ232318.1|SPN232318
[3821504]
gi|3821502|emb|AJ232317.1|SPN232317
[3821502]
gi|3821500|emb|AJ232316.1|SPN232316
[3821500]
gi|3821498|emb|AJ232315.1|SPN232315
[3821498]
gi|3821496|emb|AJ232314.1|SPN232314
[3821496]
gi|3821494|emb|AJ232313.1|SPN232313
[3821494]
gi|3821492|emb|AJ232312.1|SPN232312
[3821492]
gi|3821490|emb|AJ232311.1|SPN232311
[3821490]
gi|3821488|emb|AJ232310.1|SPN232310
[3821488]
gi|3821486|emb|AJ232309.1|SPN232309
[3821486]
gi|3821484|emb|AJ232308.1|SPN232308
[3821484]
gi|3821482|emb|AJ232307.1|SPN232307
[3821482]

gi|3821480|emb|AJ232306.1|SPN232306
[3821480]
gi|3821478|emb|AJ232305.1|SPN232305
[3821478]
gi|3821476|emb|AJ232304.1|SPN232304
[3821476]
gi|3821474|emb|AJ232303.1|SPN232303
[3821474]
gi|3821472|emb|AJ232302.1|SPN232302
[3821472]
gi|3821470|emb|AJ232301.1|SPN232301
[3821470]
gi|3821468|emb|AJ232300.1|SPN232300
[3821468]
gi|3821466|emb|AJ232299.1|SPN232299
[3821466]
gi|3821464|emb|AJ232298.1|SPN232298
[3821464]
gi|3821462|emb|AJ232297.1|SPN232297
[3821462]
gi|3821460|emb|AJ232295.1|SPN232295
[3821460]
gi|3821458|emb|AJ232294.1|SPN232294
[3821458]
gi|3821456|emb|AJ232293.1|SPN232293
[3821456]
gi|3821454|emb|AJ232292.1|SPN232292
[3821454]
gi|3821452|emb|AJ232291.1|SPN232291
[3821452]
gi|3821450|emb|AJ232290.1|SPN232290
[3821450]
gi|3821448|emb|AJ232289.1|SPN232289
[3821448]
gi|3821446|emb|AJ232288.1|SPN232288
[3821446]
gi|3821444|emb|AJ232287.1|SPN232287
[3821444]
gi|3821442|emb|AJ232286.1|SPN232286
[3821442]
gi|3821440|emb|AJ232285.1|SPN232285
[3821440]
gi|3821438|emb|AJ232284.1|SPN232284
[3821438]
gi|3821436|emb|AJ232283.1|SPN232283
[3821436]
gi|3821434|emb|AJ232282.1|SPN232282
[3821434]

gi|3821432|emb|AJ232281.1|SPN232281
[3821432]
gi|3821430|emb|AJ232280.1|SPN232280
[3821430]
gi|3821428|emb|AJ232279.1|SPN232279
[3821428]
gi|3821426|emb|AJ232278.1|SPN232278
[3821426]
gi|3821424|emb|AJ232276.1|SPN232276
[3821424]
gi|3821422|emb|AJ232275.1|SPN232275
[3821422]
gi|3821420|emb|AJ232274.1|SPN232274
[3821420]
gi|3821418|emb|AJ232273.1|SPN232273
[3821418]
gi|3821416|emb|AJ232272.1|SPN232272
[3821416]
gi|3821414|emb|AJ232271.1|SPN232271
[3821414]
gi|3821412|emb|AJ232270.1|SPN232270
[3821412]
gi|3821410|emb|AJ232269.1|SPN232269
[3821410]
gi|3821408|emb|AJ232268.1|SPN232268
[3821408]
gi|3821406|emb|AJ232267.1|SPN232267
[3821406]
gi|3821404|emb|AJ232266.1|SPN232266
[3821404]
gi|3821402|emb|AJ232265.1|SPN232265
[3821402]
gi|3821400|emb|AJ232264.1|SPN232264
[3821400]
gi|3821398|emb|AJ232263.1|SPN232263
[3821398]
gi|3821396|emb|AJ232262.1|SPN232262
[3821396]
gi|3821394|emb|AJ232261.1|SPN232261
[3821394]
gi|3821392|emb|AJ232260.1|SPN232260
[3821392]
gi|3821390|emb|AJ232259.1|SPN232259
[3821390]
gi|3821388|emb|AJ232258.1|SPN232258
[3821388]
gi|3821386|emb|AJ232257.1|SPN232257
[3821386]

gi|3821384|emb|AJ232256.1|SPN232256
[3821384]
gi|3821382|emb|AJ232255.1|SPN232255
[3821382]
gi|3821380|emb|AJ232254.1|SPN232254
[3821380]
gi|3821378|emb|AJ232253.1|SPN232253
[3821378]
gi|3821376|emb|AJ232252.1|SPN232252
[3821376]
gi|3821374|emb|AJ232251.1|SPN232251
[3821374]
gi|3821372|emb|AJ232250.1|SPN232250
[3821372]
gi|3821370|emb|AJ232249.1|SPN232249
[3821370]
gi|3821367|emb|AJ232248.1|SPN232248
[3821367]
gi|3821365|emb|AJ232247.1|SPN232247
[3821365]
gi|3821363|emb|AJ232246.1|SPN232246
[3821363]
gi|3821361|emb|AJ232245.1|SPN232245
[3821361]
gi|3821359|emb|AJ232244.1|SPN232244
[3821359]
gi|3821357|emb|AJ232243.1|SPN232243
[3821357]
gi|3821355|emb|AJ232241.1|SPN232241
[3821355]
gi|2921842|gb|AF047385.1|AF047385 [2921842]
gi|2909863|gb|AF047696.1|AF047696 [2909863]
gi|4193353|gb|AF055088.1|AF055088 [4193353]
gi|4185242|gb|AH007276.1|SEG_SPTNJUNC
[4185242]
gi|4185241|gb|AF066797.1|SPTNJUNC2
[4185241]
gi|4185240|gb|AF066796.1|SPTNJUNC1
[4185240]
gi|4097979|gb|U72655.1|SPU72655 [4097979]
gi|4063720|gb|L29323.1|STRMTR [4063720]
gi|1657605|gb|U66846.1|SPU66846 [1657605]
gi|1657602|gb|U66845.1|SPU66845 [1657602]
gi|4009485|gb|AF068903.1|AF068903 [4009485]
gi|4009477|gb|AF068902.1|AF068902 [4009477]
gi|4009462|gb|AF068901.1|AF068901 [4009462]

gi|3947767|emb|AJ233896.1|SPN233896
[3947767]
gi|3947765|emb|AJ233895.1|SPN233895
[3947765]
gi|3947763|emb|AJ233894.1|SPN233894
[3947763]
gi|3947761|emb|AJ233893.1|SPN233893
[3947761]
gi|3947759|emb|AJ233892.1|SPN233892
[3947759]
gi|3947757|emb|AJ233891.1|SPN233891
[3947757]
gi|3947755|emb|AJ233890.1|SPN233890
[3947755]
gi|3947753|emb|AJ233889.1|SPN233889
[3947753]
gi|3947751|emb|AJ233888.1|SPN233888
[3947751]
gi|3947749|emb|AJ233887.1|SPN233887
[3947749]
gi|3947730|emb|AJ233886.1|SPN233886
[3947730]
gi|3758891|emb|Z71552.1|SPADCA [3758891]
gi|3818479|gb|AF057294.1|AF057294 [3818479]
gi|2351767|gb|U89711.1|SPU89711 [2351767]
gi|3395661|dbj|AB006879.1|AB006879 [3395661]
gi|3395659|dbj|AB006878.1|AB006878 [3395659]
gi|3395657|dbj|AB006877.1|AB006877 [3395657]
gi|3395655|dbj|AB006876.1|AB006876 [3395655]
gi|3395653|dbj|AB006875.1|AB006875 [3395653]
gi|3395651|dbj|AB006874.1|AB006874 [3395651]
gi|3395649|dbj|AB006873.1|AB006873 [3395649]
gi|3395647|dbj|AB006872.1|AB006872 [3395647]
gi|3395645|dbj|AB006871.1|AB006871 [3395645]
gi|3395643|dbj|AB006870.1|AB006870 [3395643]
gi|3395641|dbj|AB006869.1|AB006869 [3395641]
gi|3395639|dbj|AB006868.1|AB006868 [3395639]
gi|2315992|gb|U87092.1|SPU87092 [2315992]
gi|2209338|gb|U93576.1|SPU93576 [2209338]
gi|2109442|gb|AF000658.1|SPDNAARG
[2109442]
gi|1881538|gb|U09239.1|SPU09239 [1881538]
gi|1666904|gb|U76218.1|SPU76218 [1666904]
gi|1613766|gb|U33315.1|SPU33315 [1613766]

gi|1498294|gb|U41735.1|SPU41735 [1498294]
gi|1213493|gb|U47687.1|SPU47687 [1213493]
gi|1163109|gb|U43526.1|SPU43526 [1163109]
gi|556001|gb|U15171.1|SPU15171 [556001]
gi|455063|gb|U02920.1|SPU02920 [455063]
gi|784896|gb|L36923.1|STRSTRH [784896]
gi|3320386|gb|AF030373.1|AF030373 [3320386]
gi|2804772|gb|AF030374.1|AF030374 [2804772]
gi|2804762|gb|AF030372.1|AF030372 [2804762]
gi|2804756|gb|AF030371.1|AF030371 [2804756]
gi|2804750|gb|AF030370.1|AF030370 [2804750]
gi|2804745|gb|AF030369.1|AF030369 [2804745]
gi|2804739|gb|AF030368.1|AF030368 [2804739]
gi|2804732|gb|AF030367.1|AF030367 [2804732]
gi|2804726|gb|AF030366.1|AF030366 [2804726]
gi|2804720|gb|AF030365.1|AF030365 [2804720]
gi|2804713|gb|AF030364.1|AF030364 [2804713]
gi|2804707|gb|AF030363.1|AF030363 [2804707]
gi|2804701|gb|AF030362.1|AF030362 [2804701]
gi|2804694|gb|AF030361.1|AF030361 [2804694]
gi|2804688|gb|AF030360.1|AF030360 [2804688]
gi|2804682|gb|AF030359.1|AF030359 [2804682]
gi|3550979|dbj|AB010387.1|AB010387 [3550979]
gi|2275100|emb|AJ000336.1|SPR6LDH [2275100]
gi|3551853|gb|AF076029.1|AF076029 [3551853]
gi|3551773|gb|U94770.1|SPU94770 [3551773]
gi|3550617|emb|AJ004869.1|SPAJ4869 [3550617]
gi|3513563|gb|AF055727.1|AF055727 [3513563]
gi|3513561|gb|AF055726.1|AF055726 [3513561]
gi|3513559|gb|AF055725.1|AF055725 [3513559]
gi|3513557|gb|AF055724.1|AF055724 [3513557]
gi|3513555|gb|AF055723.1|AF055723 [3513555]
gi|3513553|gb|AF055722.1|AF055722 [3513553]
gi|3513549|gb|AF055721.1|AF055721 [3513549]
gi|3513545|gb|AF055720.1|AF055720 [3513545]
gi|1914869|emb|Z82001.1|SPZ82001 [1914869]
gi|2911421|gb|AF046238.1|AF046238 [2911421]
gi|2911419|gb|AF046237.1|AF046237 [2911419]
gi|2911417|gb|AF046236.1|AF046236 [2911417]
gi|2911415|gb|AF046235.1|AF046235 [2911415]

gi|2911413|gb|AF046234.1|AF046234 [2911413]
 gi|2911411|gb|AF046233.1|AF046233 [2911411]
 gi|2911409|gb|AF046232.1|AF046232 [2911409]
 gi|2911407|gb|AF046231.1|AF046231 [2911407]
 gi|2911405|gb|AF046230.1|AF046230 [2911405]
 gi|3258601|gb|U40786.1|SPU40786 [3258601]
 gi|3211756|gb|AF052209.1|AF052209 [3211756]
 gi|3211752|gb|AF052208.1|AF052208 [3211752]
 gi|3211747|gb|AF052207.1|AF052207 [3211747]
 gi|3220194|gb|AF053121.1|AF053121 [3220194]
 gi|2766052|emb|Z99863.1|SPZ99863 [2766052]
 gi|2766050|emb|Z99862.1|SPZ99862 [2766050]
 gi|2766048|emb|Z99861.1|SPZ99861 [2766048]
 gi|2766046|emb|Z99860.1|SPZ99860 [2766046]
 gi|2766044|emb|Z99859.1|SPZ99859 [2766044]
 gi|2766042|emb|Z99858.1|SPZ99858 [2766042]
 gi|2766040|emb|Z99857.1|SPZ99857 [2766040]
 gi|2766038|emb|Z99856.1|SPZ99856 [2766038]
 gi|2766036|emb|Z99855.1|SPZ99855 [2766036]
 gi|2766034|emb|Z99854.1|SPZ99854 [2766034]
 gi|2766032|emb|Z99853.1|SPZ99853 [2766032]
 gi|2766030|emb|Z99852.1|SPZ99852 [2766030]
 gi|2766028|emb|Z99851.1|SPZ99851 [2766028]
 gi|2766026|emb|Z99850.1|SPZ99850 [2766026]
 gi|2766024|emb|Z99849.1|SPZ99849 [2766024]
 gi|2766022|emb|Z99848.1|SPZ99848 [2766022]
 gi|2766020|emb|Z99847.1|SPZ99847 [2766020]
 gi|2766018|emb|Z99846.1|SPZ99846 [2766018]
 gi|2766016|emb|Z99845.1|SPZ99845 [2766016]
 gi|2766014|emb|Z99844.1|SPZ99844 [2766014]
 gi|2766012|emb|Z99843.1|SPZ99843 [2766012]
 gi|2766010|emb|Z99842.1|SPZ99842 [2766010]
 gi|2766008|emb|Z99841.1|SPZ99841 [2766008]
 gi|2766006|emb|Z99840.1|SPZ99840 [2766006]
 gi|2766004|emb|Z99839.1|SPZ99839 [2766004]
 gi|2766002|emb|Z99838.1|SPZ99838 [2766002]
 gi|2766000|emb|Z99837.1|SPZ99837 [2766000]
 gi|2765998|emb|Z99828.1|SPZ99828 [2765998]
 gi|2765996|emb|Z99827.1|SPZ99827 [2765996]
 gi|2765994|emb|Z99826.1|SPZ99826 [2765994]

gi|2765992|emb|Z99825.1|SPZ99825 [2765992]
 gi|2765990|emb|Z99824.1|SPZ99824 [2765990]
 gi|2765988|emb|Z99823.1|SPZ99823 [2765988]
 gi|2765986|emb|Z99822.1|SPZ99822 [2765986]
 gi|2765984|emb|Z99821.1|SPZ99821 [2765984]
 gi|2765982|emb|Z99820.1|SPZ99820 [2765982]
 gi|2765980|emb|Z99819.1|SPZ99819 [2765980]
 gi|2765978|emb|Z99818.1|SPZ99818 [2765978]
 gi|2765976|emb|Z99817.1|SPZ99817 [2765976]
 gi|2765974|emb|Z99816.1|SPZ99816 [2765974]
 gi|2765972|emb|Z99815.1|SPZ99815 [2765972]
 gi|2765970|emb|Z99814.1|SPZ99814 [2765970]
 gi|2765968|emb|Z99813.1|SPZ99813 [2765968]
 gi|2765966|emb|Z99812.1|SPZ99812 [2765966]
 gi|2765964|emb|Z99811.1|SPZ99811 [2765964]
 gi|2765962|emb|Z99810.1|SPZ99810 [2765962]
 gi|2765960|emb|Z99809.1|SPZ99809 [2765960]
 gi|2765958|emb|Z99808.1|SPZ99808 [2765958]
 gi|2765956|emb|Z99807.1|SPZ99807 [2765956]
 gi|2765954|emb|Z99806.1|SPZ99806 [2765954]
 gi|2765952|emb|Z99805.1|SPZ99805 [2765952]
 gi|2765950|emb|Z99804.1|SPZ99804 [2765950]
 gi|2765948|emb|Z99803.1|SPZ99803 [2765948]
 gi|2894104|emb|X77249.1|SPR6CLARH [2894104]
 gi|3153897|gb|AF067128.1|AF067128 [3153897]
 gi|3152712|gb|AF065153.1|AF065153 [3152712]
 gi|3152710|gb|AF065152.1|AF065152 [3152710]
 gi|3152708|gb|AF065151.1|AF065151 [3152708]
 gi|3116426|gb|U84387.1|SPU84387 [3116426]
 gi|2385403|emb|AJ001247.1|SP7465RR3
 [2385403]
 gi|2342540|emb|AJ001250.1|SP7978RR5
 [2342540]
 gi|2342539|emb|AJ001251.1|SP7978RR3
 [2342539]
 gi|2342538|emb|AJ001248.1|SP7466RR5
 [2342538]
 gi|2342537|emb|AJ001249.1|SP7466RR3
 [2342537]
 gi|3065896|gb|AF058920.1|AF058920 [3065896]
 gi|2982647|emb|AJ002294.1|SPAJ2294 [2982647]

gi|2982645|emb|AJ002293.1|SPAJ2293 [2982645]
 gi|2982643|emb|AJ002292.1|SPAJ2292 [2982643]
 gi|2982641|emb|AJ002291.1|SPAJ2291 [2982641]
 gi|1620466|emb|X99400.1|SPDCAO [1620466]
 gi|2196665|emb|Z84381.1|HSZ84381 [2196665]
 gi|2196663|emb|Z84380.1|HSZ84380 [2196663]
 gi|2196661|emb|Z84379.1|HSZ84379 [2196661]
 gi|2196659|emb|Z84378.1|HSZ84378 [2196659]
 gi|625175|gb|L36131.1|STREXP10A [625175]
 gi|3004945|gb|AF036624.1|AF036624 [3004945]
 gi|3004943|gb|AF036623.1|AF036623 [3004943]
 gi|3004941|gb|AF036622.1|AF036622 [3004941]
 gi|3004939|gb|AF036621.1|AF036621 [3004939]
 gi|3004937|gb|AF036620.1|AF036620 [3004937]
 gi|3004935|gb|AF036619.1|AF036619 [3004935]
 gi|2370572|emb|Z86112.1|SPZ86112 [2370572]
 gi|2765946|emb|Z99802.1|SPZ99802 [2765946]
 gi|2398824|emb|Z34303.1|SPCINREC [2398824]
 gi|2894512|emb|AJ223491.1|SPPPR3 [2894512]
 gi|2198539|emb|X85787.1|SPCPS14E [2198539]
 gi|2766156|emb|Z99915.1|SPZ99915 [2766156]
 gi|2766154|emb|Z99914.1|SPZ99914 [2766154]
 gi|2766152|emb|Z99913.1|SPZ99913 [2766152]
 gi|2766150|emb|Z99912.1|SPZ99912 [2766150]
 gi|2766148|emb|Z99911.1|SPZ99911 [2766148]
 gi|2766146|emb|Z99910.1|SPZ99910 [2766146]
 gi|2766144|emb|Z99909.1|SPZ99909 [2766144]
 gi|2766142|emb|Z99908.1|SPZ99908 [2766142]
 gi|2766140|emb|Z99907.1|SPZ99907 [2766140]
 gi|2766138|emb|Z99906.1|SPZ99906 [2766138]
 gi|2766136|emb|Z99905.1|SPZ99905 [2766136]
 gi|2766134|emb|Z99904.1|SPZ99904 [2766134]
 gi|2766132|emb|Z99903.1|SPZ99903 [2766132]
 gi|2766130|emb|Z99902.1|SPZ99902 [2766130]
 gi|2766128|emb|Z99901.1|SPZ99901 [2766128]
 gi|2766126|emb|Z99900.1|SPZ99900 [2766126]
 gi|2766124|emb|Z99899.1|SPZ99899 [2766124]
 gi|2766122|emb|Z99898.1|SPZ99898 [2766122]
 gi|2766120|emb|Z99897.1|SPZ99897 [2766120]
 gi|2766118|emb|Z99896.1|SPZ99896 [2766118]

gi|2766116|emb|Z99895.1|SPZ99895 [2766116]
 gi|2766114|emb|Z99894.1|SPZ99894 [2766114]
 gi|2766112|emb|Z99893.1|SPZ99893 [2766112]
 gi|2766110|emb|Z99892.1|SPZ99892 [2766110]
 gi|2766108|emb|Z99891.1|SPZ99891 [2766108]
 gi|2766106|emb|Z99890.1|SPZ99890 [2766106]
 gi|2766104|emb|Z99889.1|SPZ99889 [2766104]
 gi|2766102|emb|Z99888.1|SPZ99888 [2766102]
 gi|2766100|emb|Z99887.1|SPZ99887 [2766100]
 gi|2766098|emb|Z99886.1|SPZ99886 [2766098]
 gi|2766096|emb|Z99885.1|SPZ99885 [2766096]
 gi|2766094|emb|Z99884.1|SPZ99884 [2766094]
 gi|2766092|emb|Z99883.1|SPZ99883 [2766092]
 gi|2766090|emb|Z99882.1|SPZ99882 [2766090]
 gi|2766088|emb|Z99881.1|SPZ99881 [2766088]
 gi|2766086|emb|Z99880.1|SPZ99880 [2766086]
 gi|2766084|emb|Z99879.1|SPZ99879 [2766084]
 gi|2766082|emb|Z99878.1|SPZ99878 [2766082]
 gi|2766080|emb|Z99877.1|SPZ99877 [2766080]
 gi|2766078|emb|Z99876.1|SPZ99876 [2766078]
 gi|2766076|emb|Z99875.1|SPZ99875 [2766076]
 gi|2766074|emb|Z99874.1|SPZ99874 [2766074]
 gi|2766072|emb|Z99873.1|SPZ99873 [2766072]
 gi|2766070|emb|Z99872.1|SPZ99872 [2766070]
 gi|2766068|emb|Z99871.1|SPZ99871 [2766068]
 gi|2766066|emb|Z99870.1|SPZ99870 [2766066]
 gi|2766064|emb|Z99869.1|SPZ99869 [2766064]
 gi|2766062|emb|Z99868.1|SPZ99868 [2766062]
 gi|2766060|emb|Z99867.1|SPZ99867 [2766060]
 gi|2766058|emb|Z99866.1|SPZ99866 [2766058]
 gi|2766056|emb|Z99865.1|SPZ99865 [2766056]
 gi|2766054|emb|Z99864.1|SPZ99864 [2766054]
 gi|2765906|emb|Z99206.1|SPZ99206 [2765906]
 gi|2765904|emb|Z99205.1|SPZ99205 [2765904]
 gi|2765902|emb|Z99204.1|SPZ99204 [2765902]
 gi|2765900|emb|Z99203.1|SPZ99203 [2765900]
 gi|2765898|emb|Z99202.1|SPZ99202 [2765898]
 gi|2765896|emb|Z99201.1|SPZ99201 [2765896]
 gi|2765894|emb|Z99200.1|SPZ99200 [2765894]
 gi|2708631|gb|AF036951.1|AF036951 [2708631]

gi|886956|emb|Z49097.1|SPCS1112X [886956]
 gi|2656093|gb|L21856.1|STRMALR [2656093]
 gi|2576332|emb|AJ002055.1|SPSPSA47 [2576332]
 gi|2576330|emb|AJ002054.1|SPSPSA2 [2576330]
 gi|2511704|emb|Y10818.1|SPY10818 [2511704]
 gi|1944619|emb|Z83335.1|SPZ83335 [1944619]
 gi|2425108|gb|AF019904.1|AF019904 [2425108]
 gi|2385404|emb|AJ001246.1|SP7465RR5 [2385404]
 gi|438213|emb|Z16082.1|PNALIB [438213]
 gi|2149613|gb|U90721.1|SPU90721 [2149613]
 gi|49391|emb|Z21841.1|SPPBP2BB [49391]
 gi|2209207|gb|AF004325.1|AF004325 [2209207]
 gi|2293061|emb|Z95914.1|SPZ95914 [2293061]
 gi|2276393|gb|U16156.1|SPU16156 [2276393]
 gi|2183314|gb|AF003930.1|AF003930 [2183314]
 gi|2182093|emb|X95717.1|SPPARECGN [2182093]
 gi|984230|emb|Z49095.1|SPCS1111A [984230]
 gi|886954|emb|Z49096.1|SPCS1092X [886954]
 gi|1181613|dbj|D82873.1|STRPBP2BE [1181613]
 gi|1181612|dbj|D82871.1|STRPBP2BCZ [1181612]
 gi|1181611|dbj|D82870.1|STRPBP2BB2 [1181611]
 gi|1181579|dbj|D82869.1|STRPBP2BA1 [1181579]
 gi|1181192|dbj|D82872.1|STRPBP2BD [1181192]
 gi|575595|dbj|D42075.1|STRPBP2B2 [575595]
 gi|1339971|dbj|D42074.1|STRPBP2B1 [1339971]
 gi|2108329|emb|Y11463.1|SPDNAGCPO [2108329]
 gi|1944115|dbj|AB002522.1|AB002522 [1944115]
 gi|1666669|emb|Z77727.1|SPIS1381C [1666669]
 gi|1666668|emb|Z77726.1|SPIS1381B [1666668]
 gi|1666667|emb|Z77725.1|SPIS1381A [1666667]
 gi|1914873|emb|Z82002.1|SPZ82002 [1914873]
 gi|1431584|emb|Z74778.1|SPDHFR [1431584]
 gi|47452|emb|Z15120.1|SPSTRG [47452]
 gi|581717|emb|Z12159.1|SPCP131G [581717]
 gi|47342|emb|X17337.1|SPAMILOC [47342]
 gi|1800300|gb|U83667.1|SPU83667 [1800300]
 gi|1532066|emb|Y07780.1|SPTETOGEN [1532066]
 gi|1161269|gb|L39074.1|STRSPXB [1161269]
 gi|1460093|emb|X94909.1|SPIGA1PRT [1460093]
 gi|1750263|gb|U72720.1|SPU72720 [1750263]
 gi|298649|gb|S56948.1|S56948 [298649]
 gi|254537|gb|S43511.1|S43511 [254537]
 gi|245227|gb|S81051.1|S81051 [245227]
 gi|245226|gb|S81045.1|S81045 [245226]
 gi|245225|gb|S81043.1|S81043 [245225]
 gi|1150618|emb|Z49988.1|SPMMSAGEN [1150618]
 gi|47456|emb|X01138.1|SPTN917A [47456]
 gi|1658316|emb|Z47210.1|SPDEXCAP [1658316]
 gi|1550802|emb|X95385.1|SPCOMCGEN [1550802]
 gi|47457|emb|X01137.1|SPTN917B [47457]
 gi|975714|emb|X90941.1|SPTRJ5251 [975714]
 gi|975713|emb|X90940.1|SPTLJ5251 [975713]
 gi|975709|emb|X90939.1|SPDNATETM [975709]
 gi|1524346|emb|Z79691.1|SOORFS [1524346]
 gi|1553054|emb|X98364.1|SPPBPHU9 [1553054]
 gi|1553052|emb|X98367.1|SPPBPHU13 [1553052]
 gi|1553050|emb|X98366.1|SPPBPHU12 [1553050]
 gi|1553048|emb|X98365.1|SPPBPHU11 [1553048]
 gi|1575029|gb|U53509.1|SPU53509 [1575029]
 gi|1542968|gb|U49088.1|SPU49088 [1542968]
 gi|1542966|gb|U49087.1|SPU49087 [1542966]
 gi|1536961|emb|Y07845.1|SPGYRA [1536961]
 gi|47391|emb|X16367.1|SPPBPX [47391]
 gi|1490398|emb|Z67739.1|SPPARCETP [1490398]
 gi|1490395|emb|Z67740.1|SPGYRBORF [1490395]
 gi|1431589|emb|Z74777.1|SPTMRDHFR [1431589]
 gi|408145|emb|Z21702.1|SPUNG MUTX [408145]
 gi|47461|emb|X61025.1|SPXISINT [47461]
 gi|47459|emb|X55651.1|SPUNGG [47459]
 gi|47454|emb|X52632.1|SPT1545E [47454]
 gi|47421|emb|Z17307.1|SPRECA [47421]
 gi|47419|emb|X67873.1|SPPONA8 [47419]
 gi|47417|emb|X67872.1|SPPONA7 [47417]
 gi|47415|emb|X67871.1|SPPONA6 [47415]

gi|47413|emb|X67870.1|SPPONA5 [47413]
 gi|47411|emb|X67869.1|SPPONA4 [47411]
 gi|47409|emb|X67867.1|SPPONA2 [47409]
 gi|47407|emb|X67866.1|SPPONA1 [47407]
 gi|47405|emb|X67868.1|SPPNA3 [47405]
 gi|47403|emb|X52474.1|SPPLY [47403]
 gi|984232|emb|X16022.1|SPPENNA [984232]
 gi|517190|emb|X78215.1|SPPBPXG [517190]
 gi|295840|emb|Z22230.1|SPPBP2BBA [295840]
 gi|288981|emb|Z22185.1|SPPBP2BAC [288981]
 gi|288979|emb|Z22184.1|SPPBP2BAB [288979]
 gi|288466|emb|Z21981.1|SPPBP2BAA [288466]
 gi|49390|emb|Z21813.1|SPPBP2XD [49390]
 gi|49389|emb|Z21812.1|SPPBP2XC [49389]
 gi|49387|emb|Z21811.1|SPPBP2BJ [49387]
 gi|49385|emb|Z21810.1|SPPBP2BI [49385]
 gi|49382|emb|Z21808.1|SPPBP2BH [49382]
 gi|49380|emb|Z21807.1|SPPBP2BG [49380]
 gi|49379|emb|Z21806.1|SPPBP2BF [49379]
 gi|49377|emb|Z21805.1|SPPBP2BE [49377]
 gi|49376|emb|Z21804.1|SPPBP2XB [49376]
 gi|49375|emb|Z21803.1|SPPBP2XA [49375]
 gi|49374|emb|Z21802.1|SPPBP2BD [49374]
 gi|49372|emb|Z21801.1|SPPBP2BC [49372]
 gi|49369|emb|Z21799.1|SPPBP2BA [49369]
 gi|47399|emb|X13137.1|SPPENASE [47399]
 gi|47397|emb|X13136.1|SPPENARE [47397]
 gi|1052802|emb|X83917.1|SPGYRBG [1052802]
 gi|587550|emb|X72967.1|SPNANA [587550]
 gi|49384|emb|Z21809.1|SPPBP1AB [49384]
 gi|49371|emb|Z21800.1|SPPBP1AA [49371]
 gi|984228|emb|Z49094.1|SPCS1091A [984228]
 gi|47372|emb|X54225.1|SPENDA [47372]
 gi|806590|emb|Z49246.1|SP667SOD [806590]
 gi|407172|emb|Z26851.1|SPATPAS2 [407172]
 gi|407166|emb|Z26850.1|SPATPAS1 [407166]
 gi|47353|emb|X63602.1|SPBOX [47353]
 gi|47348|emb|X05577.1|SPAPHA3 [47348]
 gi|47337|emb|X65132.1|SP824PBPX [47337]
 gi|47335|emb|X65134.1|SP669PBPX [47335]
 gi|47331|emb|X65133.1|SP577PBPX [47331]
 gi|559527|emb|X65136.1|SP110PBPX [559527]
 gi|311415|emb|Z22807.1|SP16SRNAA [311415]
 gi|47329|emb|X65135.1|SP531PBPX [47329]
 gi|47307|emb|X65131.1|SP290PBPX [47307]
 gi|47295|emb|X58312.1|SP16SRNA [47295]
 gi|854614|emb|Z49109.1|SPGADAGN [854614]
 gi|556428|gb|L36660.1|STORFI [556428]
 gi|511062|emb|Z35135.1|SPALLAG [511062]
 gi|1208737|gb|U47625.1|SPU47625 [1208737]
 gi|530062|gb|U12567.1|SPU12567 [530062]
 gi|153656|gb|M29686.1|STRHEXB [153656]
 gi|153654|gb|M18729.1|STRHEXA [153654]
 gi|153608|gb|M14339.1|STRDPN2A [153608]
 gi|153605|gb|M14340.1|STRDPN1A [153605]
 gi|643543|gb|U20084.1|SPU20084 [643543]
 gi|643541|gb|U20083.1|SPU20083 [643541]
 gi|643539|gb|U20082.1|SPU20082 [643539]
 gi|643537|gb|U20081.1|SPU20081 [643537]
 gi|643535|gb|U20080.1|SPU20080 [643535]
 gi|643533|gb|U20079.1|SPU20079 [643533]
 gi|643531|gb|U20078.1|SPU20078 [643531]
 gi|643529|gb|U20077.1|SPU20077 [643529]
 gi|643527|gb|U20076.1|SPU20076 [643527]
 gi|643525|gb|U20075.1|SPU20075 [643525]
 gi|643523|gb|U20074.1|SPU20074 [643523]
 gi|643521|gb|U20073.1|SPU20073 [643521]
 gi|643519|gb|U20072.1|SPU20072 [643519]
 gi|643517|gb|U20071.1|SPU20071 [643517]
 gi|643515|gb|U20070.1|SPU20070 [643515]
 gi|643513|gb|U20069.1|SPU20069 [643513]
 gi|643511|gb|U20068.1|SPU20068 [643511]
 gi|643509|gb|U20067.1|SPU20067 [643509]
 gi|1017802|gb|U37560.1|SPU37560 [1017802]
 gi|663277|gb|M36180.1|STRCOMAA [663277]
 gi|437704|gb|L20670.1|STRHYALURO [437704]
 gi|153849|gb|L07751.1|TRNTN5252R [153849]
 gi|153855|gb|M25519.1|STRVA1 [153855]
 gi|153853|gb|M80215.1|STRUVS402A [153853]
 gi|153848|gb|L07750.1|STRTN5252L [153848]

gi|153840|gb|M74122.1|STRSURPROA [153840]
gi|153796|gb|M60763.1|STRRRNAA [153796]
gi|153791|gb|M31296.1|STRRECP [153791]
gi|516639|gb|L20556.1|STRPLPA [516639]
gi|153783|gb|M28679.1|STRPROMB [153783]
gi|153782|gb|M28678.1|STRPROMA [153782]
gi|153766|gb|M90527.1|STRPONA [153766]
gi|153764|gb|J04479.1|STRPOLA [153764]
gi|153752|gb|M25515.1|STRNG4369 [153752]
gi|153722|gb|L08611.1|STRMLTODX [153722]
gi|153702|gb|J01796.1|STRMALMXP [153702]
gi|153701|gb|J01795.1|STRMALMX [153701]
gi|153693|gb|M13812.1|STRLYTPN [153693]
gi|153691|gb|M17717.1|STRLYS [153691]
gi|153667|gb|M25525.1|STRKAG73 [153667]
gi|398102|gb|L20564.1|STREXP9B [398102]
gi|398100|gb|L20563.1|STREXP9A [398100]
gi|398098|gb|L20562.1|STREXP8A [398098]
gi|398096|gb|L20561.1|STREXP7A [398096]
gi|398094|gb|L20560.1|STREXP6A [398094]
gi|398092|gb|L20559.1|STREXP5A [398092]
gi|398090|gb|L20558.1|STREXP4A [398090]
gi|153626|gb|J04234.1|STREXOA [153626]
gi|153612|gb|M11226.1|STRDPNM [153612]
gi|153603|gb|M25521.1|STRDN87669 [153603]
gi|153601|gb|M25526.1|STRDN87577 [153601]
gi|153599|gb|M25522.1|STRDN179 [153599]
gi|153594|gb|M37688.1|STRDACA [153594]
gi|153582|gb|L07752.1|STRATTB [153582]
gi|466514|gb|L31413.1|STR1RRA [466514]
gi|153551|gb|M25520.1|STR8249 [153551]
gi|153549|gb|M25524.1|STR5313972 [153549]
gi|153547|gb|M25517.1|STR29044 [153547]
gi|153545|gb|M25523.1|STR181071 [153545]
gi|153541|gb|M25518.1|STR121 [153541]
gi|153539|gb|M25516.1|STR110K70 [153539]
gi|506632|gb|U04047.1|SPU04047 [506632]
gi|393267|gb|L19055.1|STRPAPA [393267]
gi|442066|gb|S62272.1|S62272 [442066]
gi|295191|gb|L15190.1|STRPURISYN [295191]

CLAIMS

What is claimed is:

- 5 1. A method for identifying a bacteriophage coding region encoding a product active on an essential bacterial target, comprising identifying a nucleic acid sequence encoding a gene product which provides a bacteria-inhibiting function when said bacteriophage infects a host bacterium,
 wherein said bacteriophage is uncharacterized and said host bacterium
10 is a pathogenic bacterium.
2. The method of claim 1, further comprising expressing a recombinant bacteriophage ORF in cells of a bacterial strain, wherein inhibition of said cells following expression of said ORF is indicative that said product is active on an
15 essential bacterial target.
3. The method of claim 2, wherein inhibition of said bacterium following expression of said ORF is determined by comparison with the growth or viability of said bacterium following expression of an inactivated mutant form of said ORF or in
20 the absence of expression of said ORF, and wherein inhibition of said bacterium following expression of said ORF is indicative that said product is active on an essential bacterial target.
4. The method of claim 2, wherein expression of said ORF is inducible.
25
5. The method of claim 1, further comprising sequencing at least a portion of a bacteriophage genome.
6. The method of claim 1, wherein at least a portion of the nucleotide
30 sequence of a bacteriophage genome is known, said method further comprising identifying at least one ORF in said portion by computer analysis of said sequence.
7. The method of claim 6, further comprising analyzing the sequence of said at least one ORF or of a polypeptide encoded by said ORF to identify
35 homologous genes or gene products of known biochemical function, thereby indicating the biochemical function of said polypeptide.

8. The method of claim 7, wherein said homologous gene or gene product is a bacterial gene important for cell viability.

9. The method of claim 7, wherein said homologous gene or gene product is a gene or gene product known to have a bacteria-inhibiting function.

10. The method of claim 6, further comprising analyzing the sequence of said at least one ORF or of a polypeptide encoded by said ORF to identify structural motifs in said polypeptide, thereby indicating the cellular function of said polypeptide.

11. The method of claim 1, wherein a host bacterium for said bacteriophage is selected from the species group consisting of bacteria listed in Table 1.

12. The method of claim 1, wherein said bacteriophage is selected from the group consisting of uncharacterized bacteriophage listed in Table 1.

13. The method of claim 2, wherein a plurality of bacteriophage ORFs are expressed in at least one bacterium.

14. The method of claim 13, wherein each of said plurality of bacteriophage ORFs is expressed in a different bacterium.

15. The method of claim 14, wherein said plurality of bacteriophage ORFs comprises at least 10% of the ORFs in the genome of said bacteriophage.

16. The method of claim 1, wherein said pathogenic bacterium is an animal pathogen.

17. The method of claim 16, wherein said pathogenic bacterium is a human pathogen.

18. The method of claim 1, wherein said pathogenic bacterium is a plant pathogen.

19. The method of claim 1, further comprising confirming the inhibitor function of said ORF.

20. The method of claim 19, wherein said confirming comprises expressing a loss-of-function mutant form of said ORF in said host bacterium.
- 5 21. The method of claim 1, wherein said identifying a nucleic acid sequence encoding a gene product active on an essential bacterial target comprises identifying a nucleic acid sequence encoding a homolog of a bacteriophage polypeptide known to be active on an essential bacterial target.
- 10 22. The method of claim 1, wherein said identifying a bacteriophage coding region comprises identifying a first coding region from a bacteriophage having a non-pathogenic host bacterial strain related to said pathogenic bacterium, said first coding region encoding a product active on an essential bacterial target; and identifying a homolog of said first coding region, wherein said
15 homolog is a probable said bacteriophage coding region encoding a product active on an essential bacterial target.
23. The method of claim 2, wherein a plurality of bacteriophage ORFs from a plurality of different bacteriophage are expressed in at least one bacterium.
20
24. The method of claim 23, wherein each of said plurality of bacteriophage ORFs are expressed in different bacteria.
- 25 25. A method for identifying a target for antibacterial agents, comprising determining the bacterial target of an uncharacterized bacteriophage inhibitor protein.
26. The method of claim 25, wherein said determining comprises identifying at least one bacterial protein which binds to said bacteriophage inhibitor
30 protein or a fragment thereof.
27. The method of claim 26, wherein said binding is determined using affinity chromatography on a solid matrix.
- 35 28. The method of claim 25, wherein said determining comprises identifying at least one protein:protein interaction using a genetic screen.

29. The method of claim 28, wherein said genetic screen is a yeast two-hybrid screen.

30. The method of claim 25, wherein said determining comprises a co-immunoprecipitation assay or a protein-protein crosslinking assay.

31. The method of claim 25, wherein said determining comprises identifying a mutated bacterial coding sequence which protects a bacterium from said bacteriophage inhibitor.

10

32. The method of claim 25, wherein said determining comprises identifying a bacterial coding sequence which protects a bacterium against said bacteriophage inhibitor when expressed at high levels in said bacterium.

33. The method of claim 25, wherein said determining further comprises identifying a bacterial nucleic acid sequence encoding a polypeptide target of said bacteriophage inhibitor protein.

34. The method of claim 33, wherein said nucleic acid sequence is identified by determining at least a portion of the amino acid sequence of a bacterial protein target, and identifying a bacterial nucleic acid sequence which encodes said protein target.

35. The method of claim 25, wherein said bacterial target is naturally produced by a bacterial species selected from the group consisting of species of the genera listed in Table 1.

36. The method of claim 25, wherein said bacterial target is naturally produced by a bacterial strain selected from the group consisting of species listed in Table 1.

37. The method of claim 25, wherein said inhibitor protein is naturally produced by a bacteriophage selected from the group consisting of uncharacterized bacteriophage listed in Table 1.

35

38. The method of claim 25, further comprising identifying a bacteriophage ORF which encodes a product having a bacteria-inhibiting function.

39. The method of claim 38, wherein said identifying a phage ORF comprises expressing at least one bacteriophage ORF in a bacterium, wherein inhibition of said bacterium following said expression is indicative that said ORF
5 encodes a bacteria-inhibiting function.

40. The method of claim 39, wherein a plurality of bacteriophage ORFs are expressed in at least one bacterium.

10 41. The method of claim 40, wherein each of said plurality of bacteriophage ORFs is expressed in a different bacterium.

42. The method of claim 41, wherein said plurality of bacteriophage ORFs comprises at least 10% of the ORFs in the genome of said bacteriophage.
15

43. The method of claim 25, wherein said determining the bacterial target of a bacteriophage inhibitor protein is performed for a plurality of different bacteriophage of the same host bacterium.

20 44. The method of claim 25, wherein said bacterial target originates from an animal pathogen.

45. The method of claim 44, wherein said bacterial target is a gene homologous to a gene from an animal pathogen.
25

46. The method of claim 44, wherein said pathogen is a human pathogen.

47. The method of claim 25, wherein said bacterial target originates from a plant pathogen.
30

48. The method of claim 25, wherein said bacterial target is a gene homologous to a gene from a plant pathogen.

49. The method of claim 25, further comprising determining the cellular or
35 biochemical function or both of said inhibitor protein.

50. The method of claim 25, wherein said identifying the bacterial target comprises identifying a phage-specific site of action.

5 51. An isolated, purified, or enriched nucleic acid sequence at least 15 nucleotides in length, wherein said sequence corresponds to at least a portion of a bacteriophage sequence, and wherein said bacteriophage is selected from the group consisting of *Staphylococcus aureus* bacteriophage 77, 3A, 96, and 44AHJD, *Enterococcus* bacteriophage 182, and *Streptococcus pneumoniae* bacteriophage Dp-1.

10

52. The nucleic acid sequence of claim 51, wherein said sequence comprises at least 50 nucleotides.

53. The nucleic acid sequence of claim 51, wherein said nucleic acid
15 sequence corresponds to at least a portion of a nucleic acid sequence which encodes a product which provides a bacteria-inhibiting function.

54. The nucleic acid sequence of claim 53, wherein said nucleic acid
20 sequence encodes a polypeptide which provides a bacteria-inhibiting function.

25

55. The nucleic acid sequence of claim 54, wherein said nucleic acid sequence is transcriptionally linked with regulatory sequences enabling induction of expression of said sequence.

30

56. An isolated, purified, or enriched polypeptide comprising at least a portion of a protein providing a bacteria-inhibiting function, wherein said polypeptide is normally encoded by a bacteriophage selected from the group consisting of *Staphylococcus aureus* bacteriophage 77, 3A, 96, and 44AHJD, *Enterococcus* bacteriophage 182, and *Streptococcus pneumoniae* bacteriophage Dp-1.

35

57. The polypeptide of claim 56, wherein said polypeptide provides said bacteria-inhibiting function.

58. The polypeptide of claim 56, wherein said polypeptide comprises a portion at least 10 amino acid residues in length of a said polypeptide normally encoded by said bacteriophage.

59. A recombinant vector comprising a bacteriophage ORF corresponding to an ORF from a bacteriophage having a pathogenic bacterial host, wherein said
5 bacterial host is selected from the group consisting of uncharacterized bacteria of Table 1.

60. The vector of claim 59, wherein said vector is an expression vector.

10 61. The vector of claim 59, wherein said bacteriophage is selected from the group consisting of uncharacterized bacteriophage of Table 1.

62. The vector of claim 61, wherein said bacteriophage is selected from the group consisting of *Staphylococcus aureus* bacteriophage 77, 3A, 96, and 44AHJD,
15 *Enterococcus* bacteriophage 182, and *Streptococcus pneumoniae* bacteriophage Dp-1.

63. The vector of claim 60, wherein expression of said ORF is inducible.

20 64. A recombinant cell comprising a vector, wherein said vector comprises an ORF from a bacteriophage having a pathogenic bacterial host, wherein said bacterial host is selected from the group consisting of bacterial species of Table 1.

65. The recombinant cell of claim 64, wherein said bacteriophage is
25 selected from the group consisting of uncharacterized phage of Table 1.

66. The cell of claim 65, wherein said bacteriophage is selected from the group consisting of *Staphylococcus aureus* bacteriophage 77, 3A, 96, and 44AHJD,
30 *Enterococcus* bacteriophage 182, and *Streptococcus pneumoniae* bacteriophage Dp-1.

67. The cell of claim 64, wherein said vector is an expression vector and expression of said ORF is inducible.

35 68. A method for identifying an antibacterial agent, comprising identifying an active portion of a product of a bacteria-inhibiting ORF of a bacteriophage.

69. The method of claim 68, further comprising constructing a synthetic peptidomimetic molecule, wherein the structure of said molecule corresponds to the structure of said active portion.

5

70. A method for identifying a compound active on a target of a bacteriophage inhibitor protein, comprising the step of contacting a bacterial target protein with a test compound; and determining whether said compound binds to or reduces the level of activity of said target protein, wherein binding of said compound with said target protein or a reduction of the level of activity of said protein is indicative that said compound is active on said target and wherein said target is uncharacterized.

10

71. The method of claim 70, wherein said contacting is carried out *in vitro*.

15

72. The method of claim 70, wherein said contacting is carried out *in vivo* in a cell.

20

73. The method of claim 70, wherein said compound is a small molecule.

74. The method of claim 70, wherein said compound is a peptidomimetic compound.

25

75. The method of claim 70, wherein said compound is a fragment of a bacteriophage inhibitor protein.

76. The method of claim 70, further comprising determining the site of action of said compound on said target protein.

30

77. The method of claim 70, wherein said contacting is performed for a plurality of said target proteins.

35

78. A method of screening for potential antibacterial agents, comprising the step of determining whether any of a plurality of compounds is active on a target of a bacteriophage inhibitor protein,

wherein said target is naturally produced by a pathogenic bacterium.

79. The method of claim 78, wherein said plurality of compounds are small molecules.

5

80. The method of claim 78, wherein said determining is performed for a plurality of said targets.

10

81. A method for inhibiting a bacterium, comprising the step of; contacting said bacterium with a compound active on a target of a bacteriophage inhibitor protein, wherein said target or the target site is uncharacterized.

15

82. The method of claim 81, wherein said compound is said protein or an active fragment thereof.

83. The method of claim 81, wherein said compound is a structural mimetic of said protein.

20

84. The method of claim 81, wherein said compound is a small molecule.

85. The method of claim 81, wherein said contacting is performed *in vitro*.

25

86. The method of claim 81, wherein said contacting is performed *in vivo* in an animal.

87. The method of claim 86, wherein said animal is a human.

30

88. The method of claim 81, wherein said contacting is carried out *in vivo* in a plant.

89. The method of claim 81, wherein said bacterium is selected from the group of bacteria listed in Table 1.

35

90. A method for treating a bacterial infection in an animal suffering from an infection, comprising administering to said animal a therapeutically effective amount of compound active on a target of a bacteriophage inhibitor protein in a bacterium involved in said infection,

5 wherein said target is an uncharacterized target or the compound is active at an uncharacterized target site.

91. The method of claim 90, wherein said compound is a small molecule.

10 92. The method of claim 90, wherein said compound is a peptidomimetic compound.

93. The method of claim 90, wherein said compound is a fragment of a bacteriophage inhibitor protein.

15

94. The method of claim 90, wherein said animal is a mammal.

95. The method of claim 94, wherein said mammal is a human.

20 96. The method of claim 90, wherein said bacterium is selected from the group listed in Table 1.

97. The method of claim 90, wherein said bacteriophage inhibitor protein is from a bacteriophage selected from the group of bacteriophage listed in Table 1.

25

98. A method for prophylactically treating an animal at risk of an infection, comprising administering to said animal a prophylactically effective amount of a compound active on a target of a bacteriophage inhibitor protein,

30 wherein said target is an uncharacterized target or the site of action of said compound is an uncharacterized target site.

99. The method of claim 98, wherein said compound is a small molecule.

35 100. The method of claim 98, wherein said compound is a peptidomimetic compound.

101. The method of claim 98, wherein said compound is a fragment of a bacteriophage inhibitor protein.

102. The method of claim 98, wherein said animal is a mammal.

103. The method of claim 102, wherein said mammal is a human.

104. An antibacterial agent active on a target of a bacteriophage inhibitor protein, wherein said target is an uncharacterized target or said agent is active at a phage-specific site on said target.

105. The agent of claim 104, wherein said agent is a peptidomimetic of a bacteriophage inhibitor polypeptide.

106. The agent of claim 104, wherein said agent is a small molecule.

107. The agent of claim 104, wherein said agent is a fragment of a bacteriophage inhibitor polypeptide.

108. The agent of claim 104, wherein said agent is active at a phage-specific site on said target.

109. A method of making an antibacterial agent, comprising the steps of:
a) identifying a target of a bacteriophage inhibitor polypeptide;
b) screening a plurality of test compounds to identify a compound active on said target; and
c) synthesizing said compound in an amount sufficient to provide a therapeutic effect when administered to an organism infected by a bacterium naturally producing said target.

110. The method of claim 109, wherein said compound is a small molecule.

111. The method of claim 109, wherein said compound is a peptidomimetic compound.

112. The method of claim 109, wherein said compound is a fragment or derivative of a bacteriophage inhibitor protein.

5 113. A computer readable device having recorded therein a nucleotide sequence of a portion of at least one bacteriophage genome of *Staphylococcus aureus* bacteriophage 77, bacteriophage 3A, or bacteriophage 96, a nucleotide sequence at least 95% identical to a said nucleotide sequence, a ribonucleic acid equivalent, a degenerate equivalent, a homologous sequence, or at least one amino acid sequence
10 encoded by said nucleotide sequence; and
a nucleotide sequence or amino acid sequence analysis program,
wherein said program can perform at least one sequence analysis on said nucleotide or amino acid sequence.

15 114. The device of claim 113, wherein said at least a portion of at least one bacteriophage genome comprises at least one ORF.

 115. The device of claim 113, wherein said device comprises a medium selected from the group consisting of floppy disk, computer hard drive, optical disk,
20 computer random access memory, and magnetic tape wherein said nucleotide or amino acid sequence or said program or both are recorded on said medium.

 116. The device of claim 113, wherein said portion of at least one bacteriophage genomic nucleotide sequence comprises at least 50% of at least one
25 bacteriophage genomic sequence.

 117. The device of claim 113, wherein said at least one bacteriophage nucleotide genomic sequence comprises portions of a plurality of bacteriophage nucleotide genomic sequences.
30

 118. A computer-based system for identifying biologically important portions of a bacteriophage genome, comprising:
a) a data storage medium having recorded thereon a nucleotide sequence
35 corresponding to a portion of at least one bacteriophage genome, wherein said bacteriophage genome is uncharacterized;

- b) a set of instructions allowing searching of said sequence to analyze said sequence; and
- c) an output device.

5 119. The system of claim 118, wherein said output device comprises comprises a device selected from the group consisting of a printer, a video display, and a recording medium.

10 120. The system of claim 118, wherein said bacteriophage genome is of a bacteriophage selected from the group consisting of uncharacterized bacteriophage listed in Table 1.

15 121. The system of claim 118, wherein said uncharacterized bacteriophage is selected from the group consisting of bacteriophage 77, 3A, and 96.

 122. A method for identifying or characterizing a bacteriophage ORF, comprising the steps of:

- a) providing a computer-based system for analyzing nucleic acid or
20 amino acid sequence data, wherein said system comprises a data storage medium having recorded thereon at least one nucleotide or amino acid sequence corresponding to a portion of at least one uncharacterized bacteriophage genome, a set of instructions allowing searching of said sequence to analyze said sequence; and an output device;
- b) analyzing at least a portion of at least one said sequence; and
- 25 c) outputting results of said analyzing to said output device.

 123. The method of claim 122, wherein said analysis identifies sequence similarity or homology with sequences selected from the group consisting of bacterial ORFs encoding products with related biological function; ORFs encoding known
30 inhibitors or bacteria, essential bacterial ORFs.

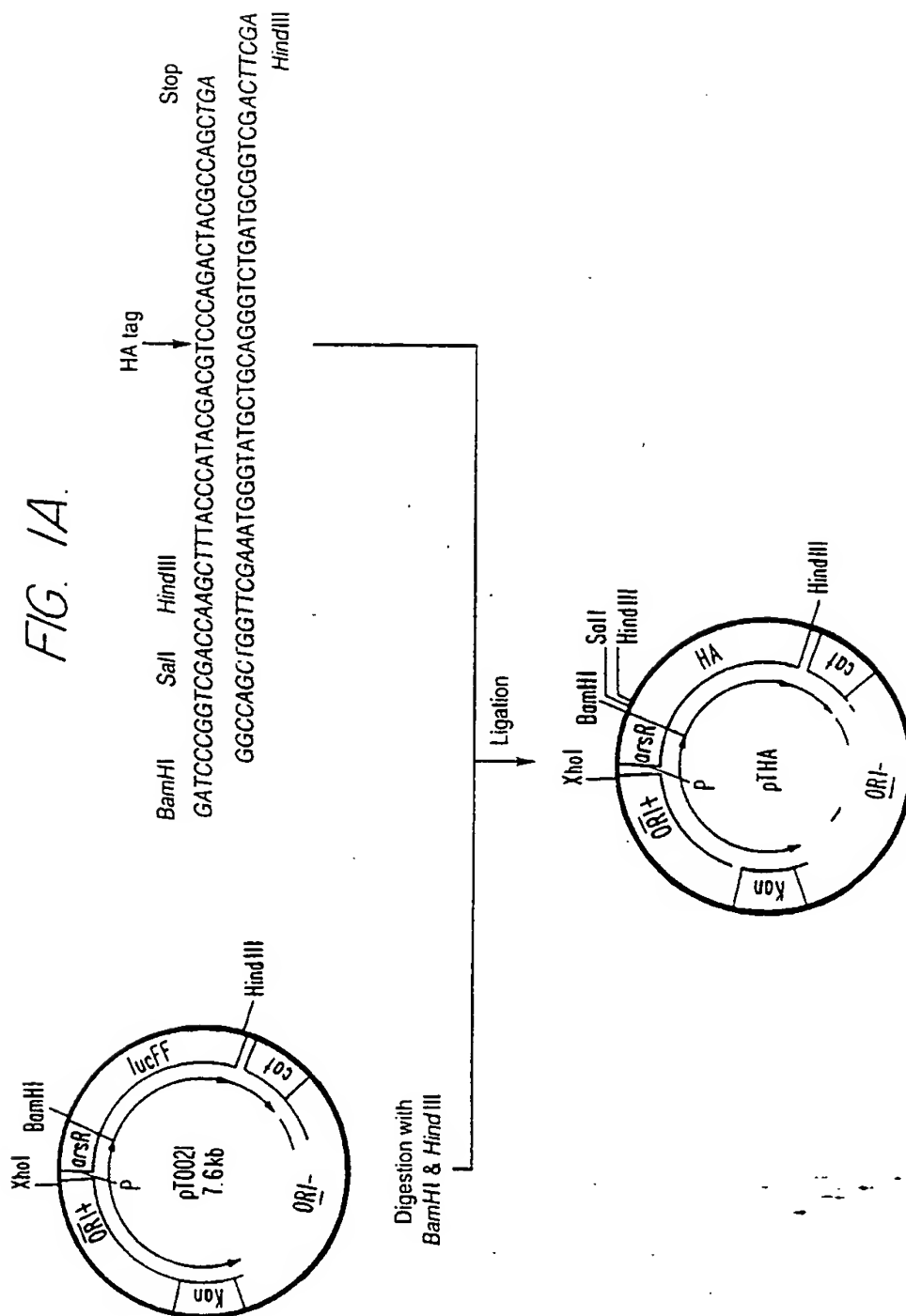
 124. The method of claim 122, wherein said analysis comprises identifying a probable biological function based on identification of structural elements or sequence homology or similarity.

35 125. The method of claim 122, wherein said bacteriophage is selected from the group consisting of uncharacterized bacteriophage listed in Table 1.

126. The method of claim 125, wherein said uncharacterized bacteriophage is selected from bacteriophage 77, 3A, and 96.

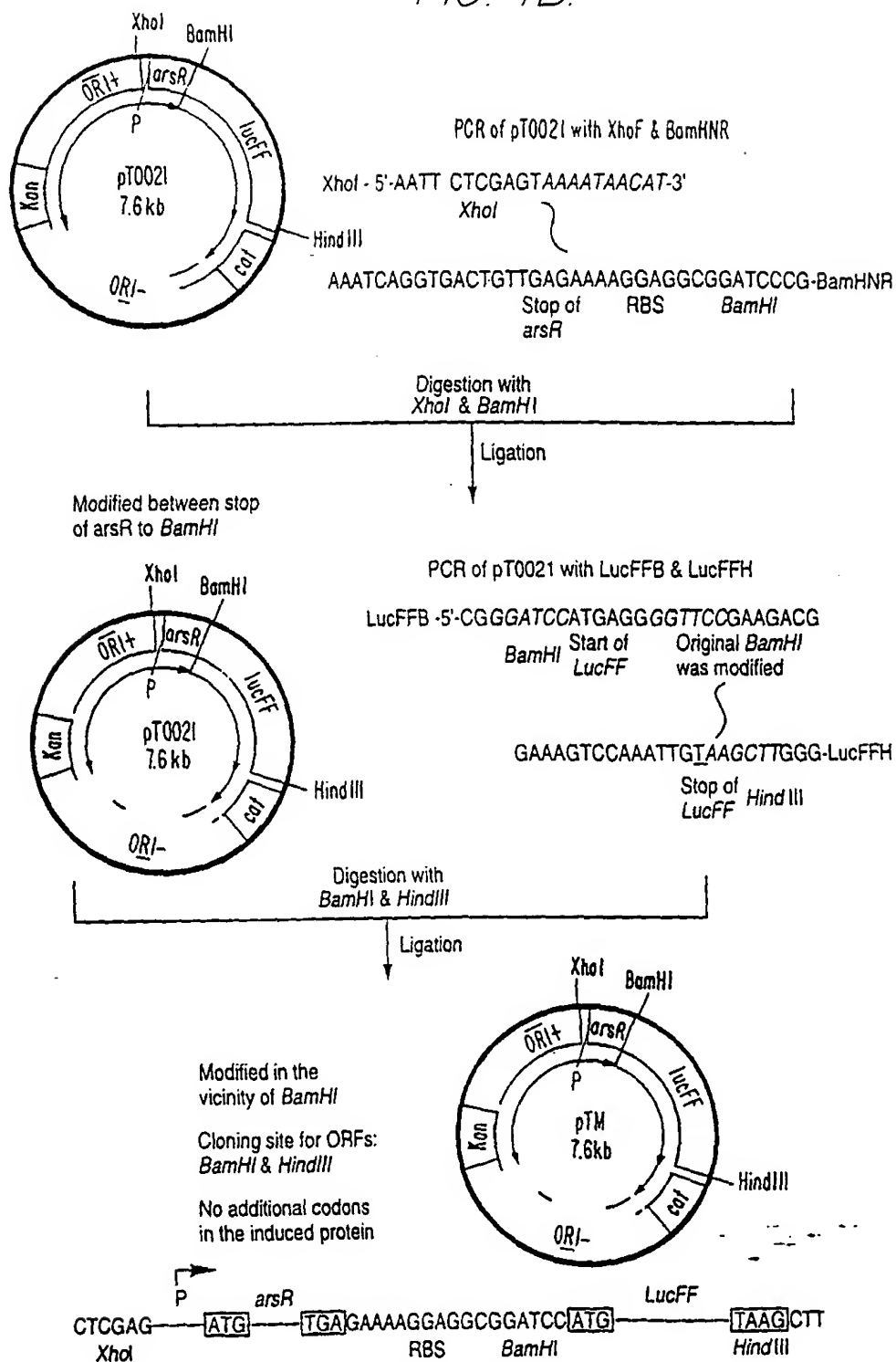
01/11

FIG. 1A.



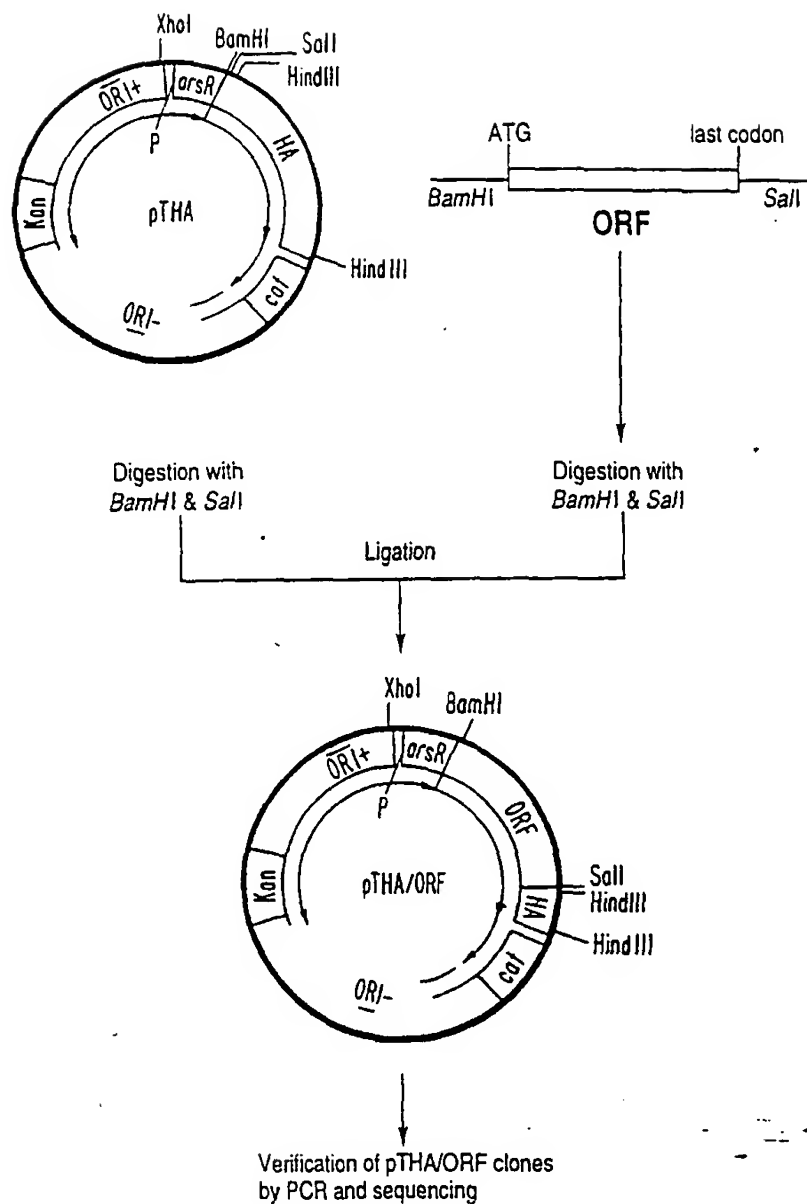
02/11

FIG. 1B.



03/11

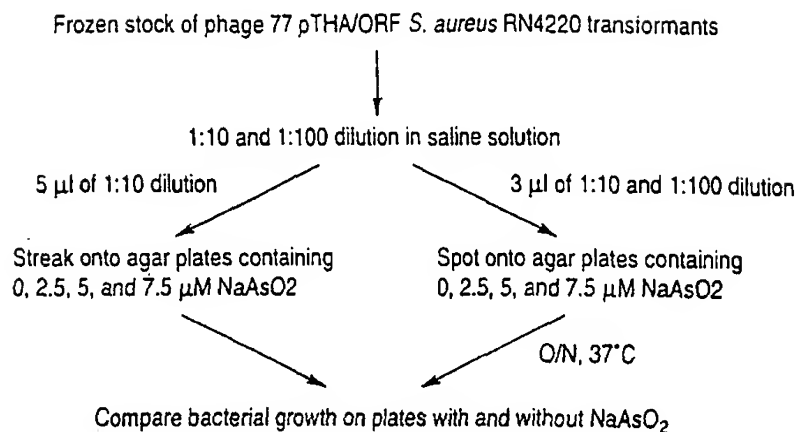
FIG. 2.



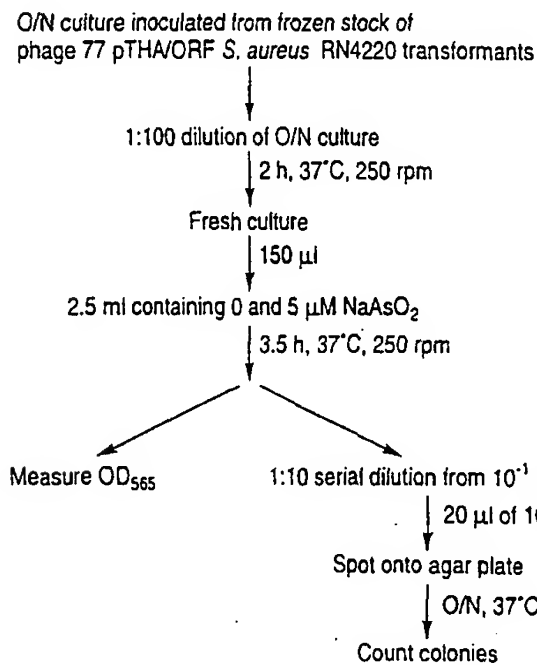
04/11

FIG. 3.

(A) Functional assay on semi-solid support media

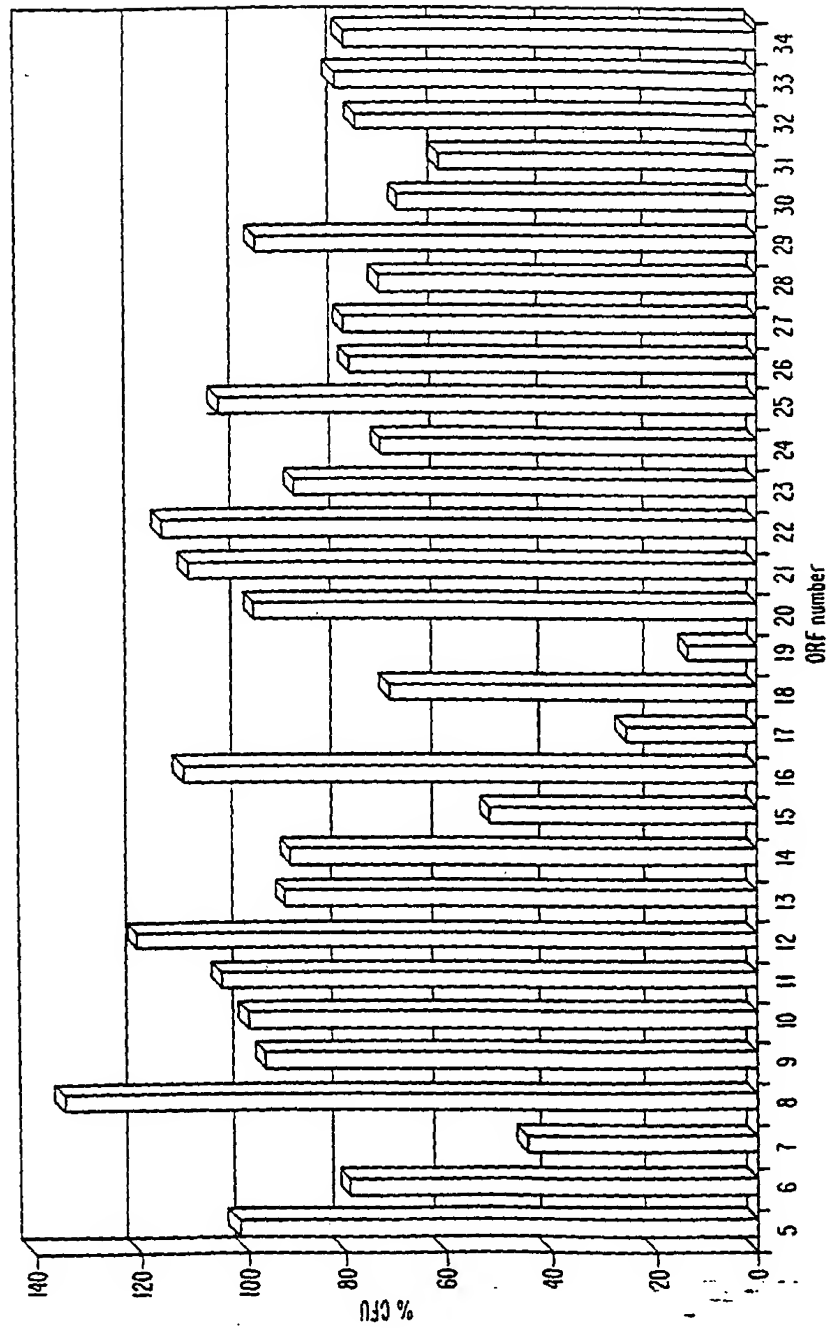


(B) Functional assay in liquid medium



SUBSTITUTE SHEET (RULE 26)

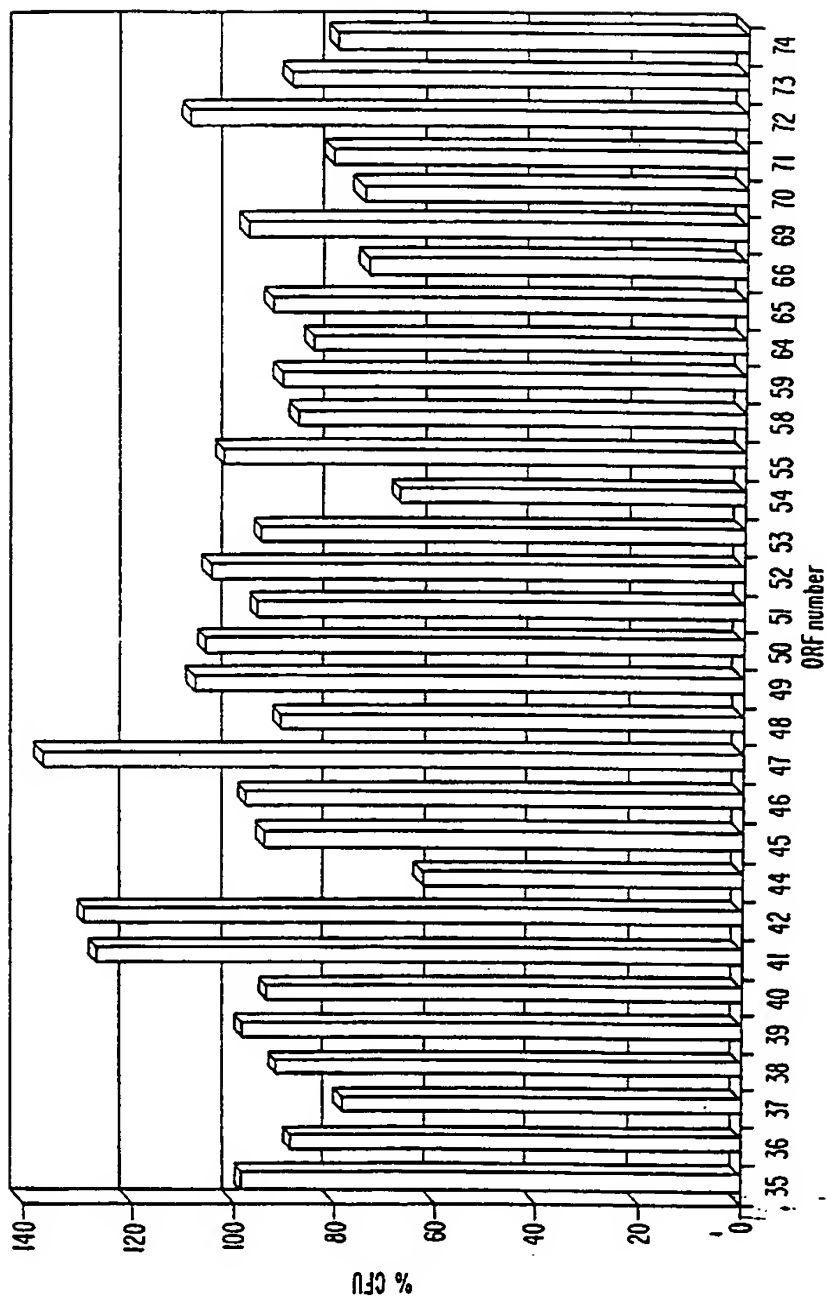
05/11

A. Inhibition of bacterial growth with individual ORFs of a *S. aureus* Bacteriophage.

06/11

B. Inhibition of bacterial growth with individual ORFs of a *S. aureus* Bacteriophage.

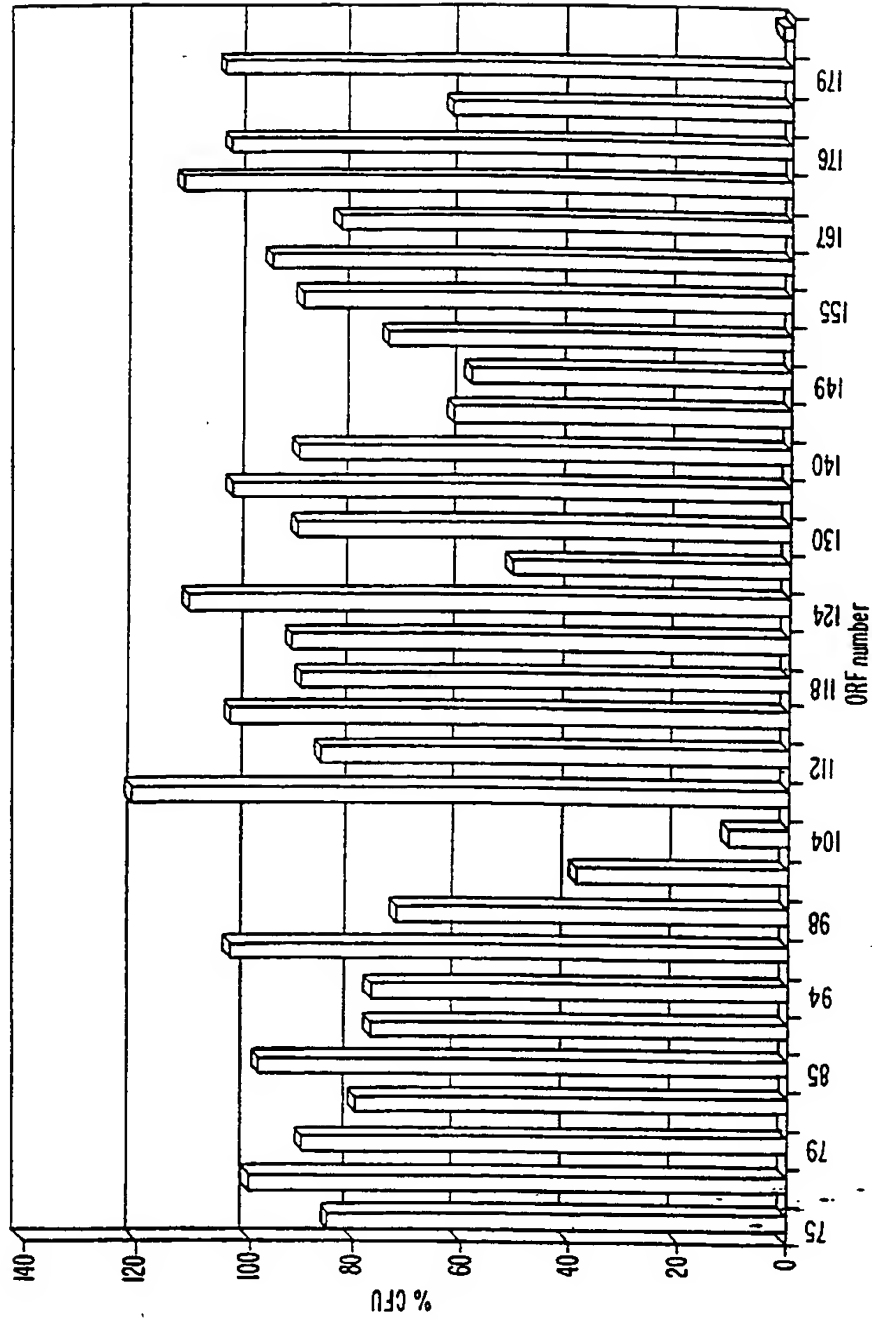
FIG. 4B.



07/11

C. Inhibition of bacterial growth with individual ORFs of a *S. aureus* Bacteriophage.

FIG. 4C.



08/11

FIG. 5.

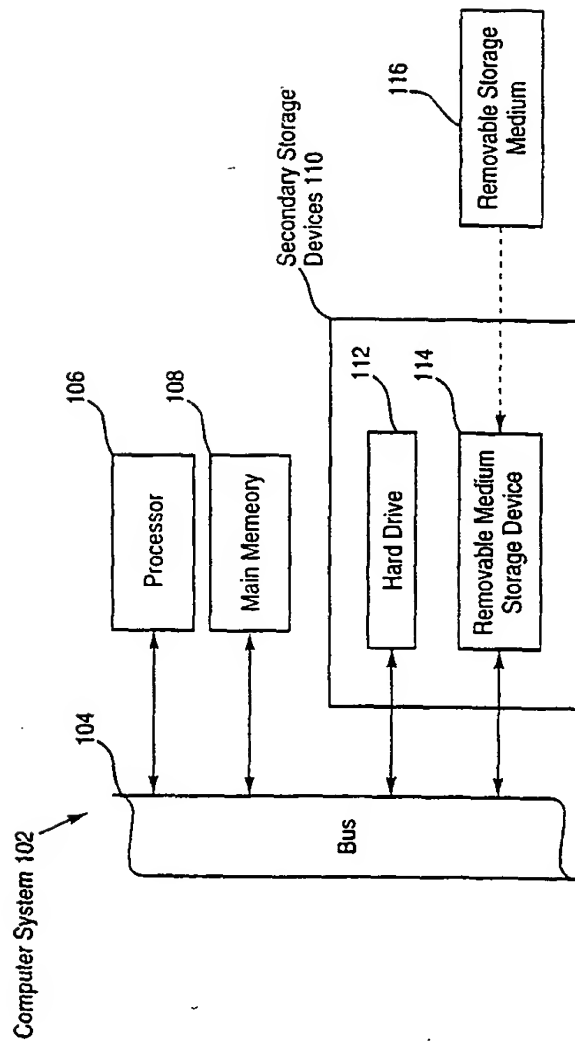
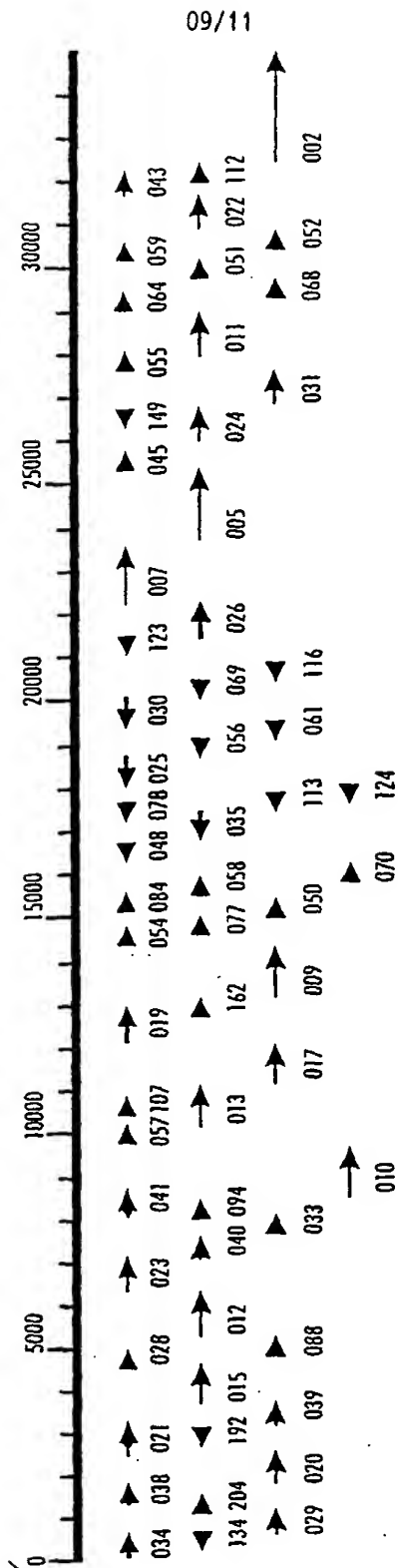


Fig 6A Fig 6B

Fig. 6

Fig. 6A

PHAGE: BACTERIOPHAGE Dp1
 MINIMAL ORF SIZE: 33 A. A.
 ORFS "WITH" RBS.
 NUMBER OF ORFS: 85



029: exsB; 012: DNA pol. III beta;

038: exsC; 6-PYRUVYLTYLHYDROPTERIN

020: exsD; COENZYME PQ; 013: DNA POL III GAMMA AND TAU;

021: GTP CYCLOHYDROLASE;

039: CITRULLINE BIOSYNTHESIS;

041: dUTPASE;

010: RecA;

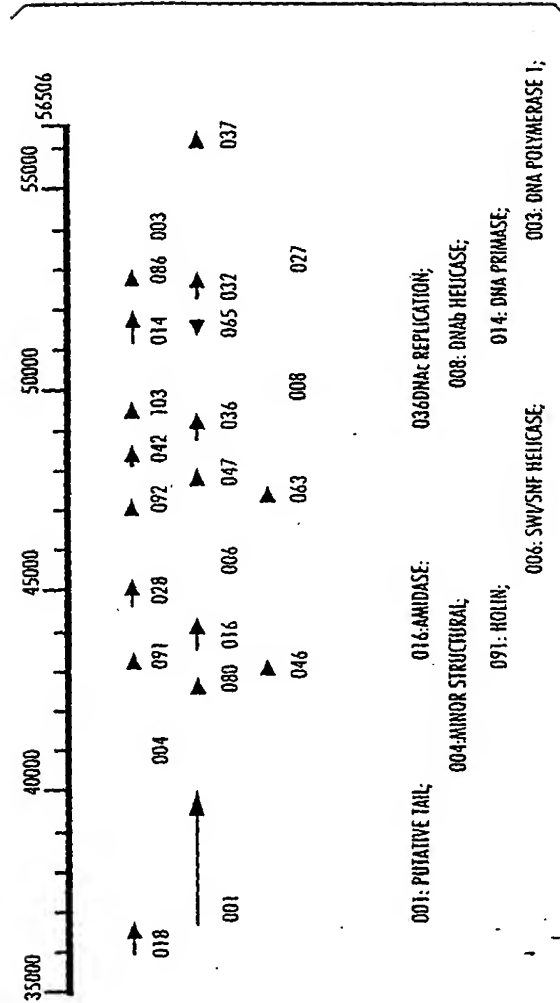
007: TERMINASE;

011: MAJOR HEAD;

002: TAIL;

10/11

Fig. 6B



11/11

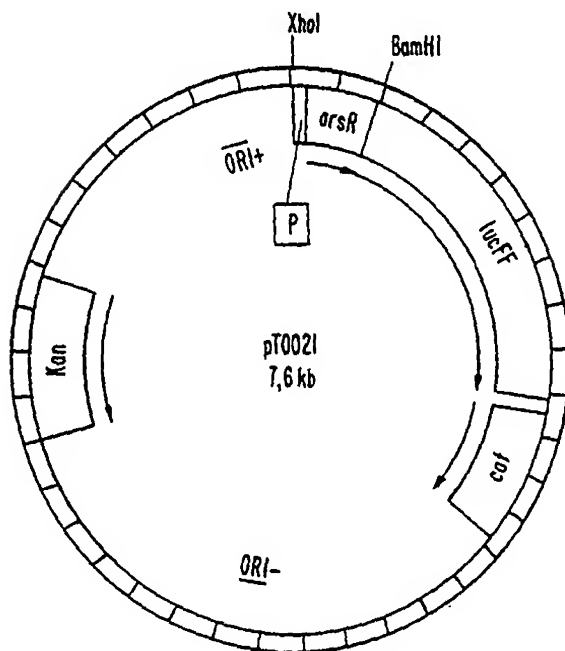
FIG. 7.

Abbreviations:

kan: gene encoding kanamycin resistance
 cat: gene encoding chloramphenicol resistance
 ori + and -: origin of replication in gram-positive and
 gram-negative bacteria, respectively
 arsR: gene encoding regulatory protein of the ars promoter
 P: ars promoter
 lucFF: gene encoding luciferase protein. This portion will
 be removed and replaced by individual *S. aureus* phage
 genes.

Reference:

Tauriainen et al., Appl. Environ. Microbio. 1997. 63: 4456-
 4461



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 June 2000 (08.06.2000)

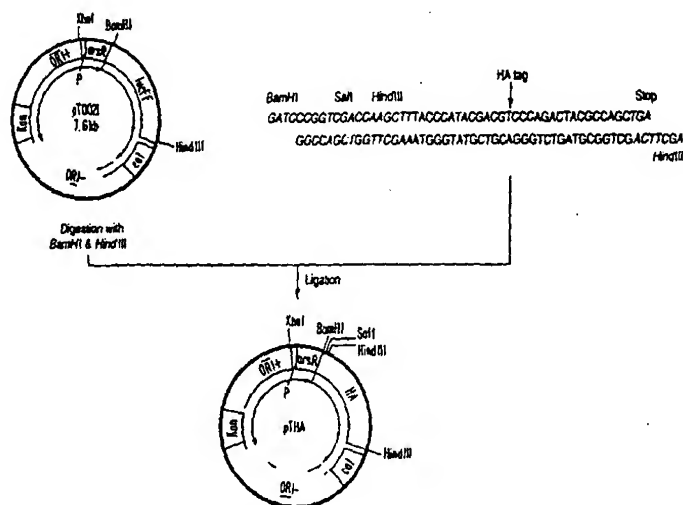
PCT

(10) International Publication Number
WO 00/32825 A3

- (51) International Patent Classification⁷: C12Q 1/70, 1/68, C12N 15/10, 15/34, 1/21, C07K 14/01, C12Q 1/18
- (21) International Application Number: PCT/IB99/02040
- (22) International Filing Date: 3 December 1999 (03.12.1999)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/110,992 3 December 1998 (03.12.1998) US
09/326,144 3 June 1999 (03.06.1999) US
09/407,804 28 September 1999 (28.09.1999) US
60/157,218 30 September 1999 (30.09.1999) US
60/168,777 1 December 1999 (01.12.1999) US
09/454,252 2 December 1999 (02.12.1999) US
- (71) Applicant (for all designated States except US): PHAGETECH, INC. [CA/CA]; Place du Parc, Case Postale 387, Montreal H2W 2N9 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): PELLETIER, Jerry [CA/CA]; 8 Lakeview, Baie D'Urfe, Quebec H9X 3B1 (CA). GROS, Phillippe [CA/CA]; 107 Montrose, St. Lambert, Quebec J4R 1X4 (CA). DUBOW, Michael [CA/CA]; 4901 Coolbrook Avenue, Montreal, Quebec H3X 2K8 (CA).
- (74) Agents: MORROW, Joy, D. et al.; Smart & Biggar, 900 - 55 Metcalfe Street, P.O. Box 2999, Station D, Ottawa, Ontario K1P 5Y6 (CA).
- (81) Designated States (national): AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: DEVELOPMENT OF ANTI-MICROBIAL AGENTS BASED ON BACTERIOPHAGE GENOMICS



WO 00/32825 A3

(57) Abstract: A method for identifying suitable targets for antibacterial agents based on identifying targets of bacteriophage-encoded proteins is described. Also described are compositions useful in the identification methods and in inhibiting bacterial growth, and methods for preparing and using such compositions.



Published:

— *With international search report.*


For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(88) Date of publication of the international search report:

18 January 2001

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/06718

A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : A01N 63/00; A61K 39/02 US CL : 424/93.6, 234.1 According to International Patent Classification (IPC) or to both national classification and IPC																				
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 424/93.6, 234.1 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) WEST, DIALOG, MEDLINE, BIOSIS, SCISEARCH, EMBASE search terms; vancomycin resistant enterococcus faecium, phage, bacteriophage, lytic, ENB6, ENB13, treating, antibiotic																				
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>US 5,688,501 A (MERRIL et al) 18 November 1997, see entire document.</td> <td>1-13</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	A	US 5,688,501 A (MERRIL et al) 18 November 1997, see entire document.	1-13												
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.																		
A	US 5,688,501 A (MERRIL et al) 18 November 1997, see entire document.	1-13																		
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.																				
<table border="0"> <tr> <td>* Special categories of cited documents:</td> <td>*T*</td> <td>later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</td> </tr> <tr> <td>*A* document defining the general state of the art which is not considered to be of particular relevance</td> <td>*X*</td> <td>document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</td> </tr> <tr> <td>*E* earlier document published on or after the international filing date</td> <td>*Y*</td> <td>document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</td> </tr> <tr> <td>*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td> <td>*A*</td> <td>document member of the same patent family</td> </tr> <tr> <td>*O* document referring to an oral disclosure, use, exhibition or other means</td> <td></td> <td></td> </tr> <tr> <td>*P* document published prior to the international filing date but later than the priority date claimed</td> <td></td> <td></td> </tr> </table>			* Special categories of cited documents:	*T*	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	*A* document defining the general state of the art which is not considered to be of particular relevance	*X*	document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	*E* earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A*	document member of the same patent family	*O* document referring to an oral disclosure, use, exhibition or other means			*P* document published prior to the international filing date but later than the priority date claimed		
* Special categories of cited documents:	*T*	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention																		
A document defining the general state of the art which is not considered to be of particular relevance	*X*	document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone																		
E earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art																		
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A*	document member of the same patent family																		
O document referring to an oral disclosure, use, exhibition or other means																				
P document published prior to the international filing date but later than the priority date claimed																				
Date of the actual completion of the international search 26 JUNE 2000		Date of mailing of the international search report 03 AUG 2000																		
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230		Authorized officer BRETT L NELSON  Telephone No. (703) 308-0196																		